Analysis of the influence of patients setup errors on the doses received by them during radiotherapy

Analiza wpływu niepewności ułożenia pacjentów na dawki otrzymane przez nich podczas radioterapii.

The Doctoral Thesis

Author:

Marta Giżyńska

Supervisors:

dr hab. Paweł Kukołowicz, prof. nadzw. COI
(Maria Sklodowska-Curie Institute — Oncology Center)

dr hab. Maciej Kamiński
(University of Warsaw, Faculty of Physics)

Warsaw, June 2019

Acknowledgements

First, I would like to thank Paweł Kukołowicz and Maciej Kamiński for supervising me during my PhD studies and research.

Presented work would not be possible without patients' datasets. I am grateful to Maria Piziorska from Maria Sklodowska-Curie Institute — Oncology Center and Maarten L.P. Dirkx and Andras G. Zolnay from Erasmus MC for providing the prostate setup errors from their radiotherapy departments databases.

I would like to acknowledge my colleagues from Maria Sklodowska-Curie Institute — Oncology Center: Dorota Blatkiewicz, Beata Czyżew, Małgorzata Gil-Ulkowska, Sandra Łukomska-Maluszczak, Anna Paciorkiewicz, Magdalena Ziemek as well as my former students from Warsaw University: Adam Cichoński, Maciej Gałecki and Anna Zaleska for their enthusiastic collaboration. Without their help it would be very difficult to perform the study concerning estimation of cumulative dose.

I am grateful to professor Zbigniew Szutkowski, head of radiotherapy department at Maria Sklodowska-Curie Institute — Oncology Center, for allowing me to present photographs taken in the department. I appreciate very much eager permission of patients who allowed me to take the pictures.

I am also thankful to Jarosław Żygierewicz and Maciej Kamiński from University of Warsaw as well as Krzysztof Postek from Erasmus University for their support in statistical issues.

My special gratefulness goes to professor Ben Heijmen from Erasmus MC for his great support, encouragement, enthusiastic discussions and constructive comments. I appreciate very much all the time you dedicated to our meetings and collaboration. This project would not be the same without it.

Last but not least I would like to thank my family and friends who were supporting me during that time.

Marta Giżyńska

Abstract

Radiotherapy is currently one of the most important treatment methods used in cancer treatment. Its aim is to deliver prescribed dose to the target volume while sparing healthy tissues. In order to reach this aim radiotherapy treatment should be applied in as precise manner as possible. Therefore patients setup uncertainties and their influence on the dose received during radiotherapy course was a subject of many research done through last decades. Studies at that time based on limited amount of information. Nowadays, technological progress in radiotherapy treatment allows to collect setup data during entire course of treatment. Analysis of such a big data are still lacking.

In presented work two sets of clinical data were analyzed. First group consisted of 100 patients treated in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw (Poland), second group consisted of 835 patients treated in Erasmus MC University Medical Center Rotterdam (Netherlands). Both groups included data of setup errors for prostate cancer patients. The difference was in number of fractions received during entire treatment and methods used to detect setup errors (bony anatomy or gold markers).

The statistical analysis of the data collected for both populations showed that they were not normally distributed. Presence of inter-fraction time trends was investigated as one of possible reasons of non-normal distribution of setup errors. As in limited fraction treatment some trends can always be detected, the method to distinguish physiological trends (i.e. related to patient physiology and anatomy) from these which can be attributed to limited number of measurements was proposed.

The presented studies were done on two previously mentioned clinical datasets as well as on many different datasets with artificial setup errors mimicking the natural patients behavior — so called synthetic populations. Those synthetic populations were created in order to cover wide range of clinically relevant population parameters. Current methods of dealing with setup errors and incorporating this knowledge into the treatment planning and delivery process are mostly based on population parametrizations and concept of cumulative dose. In this work a parametrization taking into account time trends was investigated and compared with the conventional one, most often used in the clinical practice. It was shown that a wrong parametrization used to describe population of patients setup errors may lead to the underestimation of CTV-PTV margins. The investigated parametrization may be further used in margin-less planning.

In the presented work, apart from investigating alternative parametrization method, a possibility of estimating patient specific cumulative dose was examined. That methodology can be further used for individualization of margin-less planning algorithms.

One of the methods to deal with setup errors and uncertainties is to apply off-line and on-line verification protocols. Offline verification protocols require less workload. Still it is known from clinical practice that some patients will require an on-line verification. The mostly applied

No Action Level protocol efficiency is related to mean setup error estimation. A modification of this protocol was proposed in order to provide an easy method to distinguish patients who require more attention (i.e. more fractions with imaging). The modified protocol does not require additional workload in terms of average number of fractions with imaging. Instead of doing imaging in the same number of fractions for each patient, the fractions with imaging are not-equally spread within the population of patients.

Although off-line verification protocols may reduce the systematic errors, some residual errors will always be present. To ensure a proper irradiation of the target some margin is added to it during the treatment planning process. Currently the most applied margin recipes are based on conventional parametrization. That is why it does not take into account inter-fraction time trends. As existence of this trends was shown during population data analysis, the new recipe for margin calculation was provided. In limit of no time trends this new recipe simplifies to currently used one.

Keywords:

radiotherapy, setup errors, cumulative dose, verification protocols, CTV-PTV margins, time trends, Monte Carlo simulations

Streszczenie

Radioterapia jest obecnie jedną z najważniejszych metod leczenia nowotworów. Jej celem jest dostarczenie przepisanej dawki do obszaru tarczowego przy jednoczesnym zminimalizowaniu dawki w tkankach zdrowych. Aby osiągnąć ten cel radioterapia powinna być przeprowadzana jak najbardziej precyzyjnie. Z tego powodu niepewności ułożenia pacjentów i ich wpływ na dawki dostarczone w trakcie kursu radioterapii były przedmiotem wielu badań w czasie minionych dekad. Ówczesne badania bazowały na ograniczonej liczbie danych. Postęp technologiczny w radioterapii umożliwia obecnie zbieranie danych dotyczących niepewności ułożenia pacjenta podczas całego procesu leczenia. Ciągle jednak nieliczne są badania analizujące takie dane.

W prezentowanej pracy przeprowadzono analizę dwóch zbiorów danych klinicznych. Pierwszą grupę stanowiło 100 pacjentów leczonych w Centrum Onkologii — Instytucie im. Marii Skłodowskiej-Curie w Warszawie (Polska), drugą grupę stanowiło 835 pacjentów leczonych w szpitalu uniwersyteckim Erasmus MC w Rotterdamie (Holandia). Obie grupy zawierały dane dotyczące niepewności ułożenia pacjentów leczonych z powodu nowotworu prostaty. Grupy różniły się liczbą frakcji radioterapii jaka składała się na kurs leczenia a także metodami użytymi do oceny niepewności ułożenia (struktury kostne i znaczniki złota).

Statystyczna analiza danych zebranych dla obu populacji pokazała, iż nie zawsze podlegają one rozkładowi normalnemu. Obecność trendów czasowych pomiędzy frakcjami leczenia była badana jako jedna z potencjalnych przyczyn znalezionych rozbieżności w stosunku do rozkładów normalnych. w przypadku skończonej liczby frakcji pewien trend czasowy może być zawsze znaleziony, dlatego opracowano metodę rozróżniania trendów fizjologicznych (tj. związanych z fizjologią i anatomią pacjenta) od tych, które mogą być wyjaśnione skończoną liczbą pomiarów. Prezentowane w pracy badania przeprowadzono na dwóch, wspomnianych wcześniej, klinicznych bazach danych niepewności ułożeń pacjentów, a także na licznych innych bazach danych ze sztucznie generowanymi niepewnościami ułożenia pacjentów odzwierciedlającymi ich naturalne zachowanie — tzw. populacjach syntetycznych. Te syntetyczne populacje zostały wygenerowane w taki sposób by pokryć jak największy zakres klinicznie istotnych parametrów.

Obecne metody uwzględniania niepewności ułożenia pacjentów w procesie planowania i realizacji radioterapii bazują przede wszystkim na parametryzacji populacji pacjentów i koncepcji dawki skumulowanej. W prezentowanej pracy parametryzacja populacji uwzględniająca trendy czasowe została przebadana i porównana z konwencjonalną parametryzacją, najczęściej wykorzystywaną w praktyce klinicznej. Zostało pokazane, że niewłaściwy dobór parametryzacji użytej do opisu populacji niepewności ułożenia pacjentów może prowadzić do zbyt małych marginesów CTV-PTV. W prezentowanej pracy, poza zaproponowaniem alternatywnej parametryzacji, zbadano możliwości estymowania skumulowanej dawki dla pojedynczego pacjenta. Zaproponowana metodologia, a w szczególności badana parametryzacja, może zostać w przyszłości użyta w zindywidualizowanej wersji algorytmów planowania bez marginesów.

Jedną z metod minimalizowania niepewności ułożenia pacjentów jest stosowanie protokołów

weryfikacji off-line i on-line. Protokoły weryfikacji off-line wymagają mniejszego nakładu pracy. Wiadomo jednak z praktyki klinicznej, że niektórzy pacjenci będą wymagali zastosowania weryfikacji on-line. Szeroko stosowany protokół weryfikacji „No Action Level" opiera się na estymacie średniej niepewności ułożenia pacjenta. W niniejszej pracy została zaproponowana modyfikacja tego protokołu umożliwiająca rozróżnienie pacjentów wymagających większej uwagi (tj. większej liczby frakcji z obrazowaniem). Zmodyfikowany protokół nie wymaga zwiększonego nakładu pracy rozumianego jako średnia ilość frakcji z obrazowaniem w populacji pacjentów. Wykonywanie obrazowania w tej samej liczbie frakcji u wszystkich pacjentów zastąpiono w proponowanej metodzie nierównomierną dystrybucją frakcji z obrazowaniem w populacji pacjentów.

Protokoły weryfikacji off-line mają za zadanie zmniejszenie niepewności systematycznych jednak pewne niepewności rezydualne zawsze będą obecne. Aby zapewnić prawidłowe napromienienie obszaru tarczowego pewien margines jest dodawany do niego w trakcie procesu planowania radioterapii. Obecnie najczęściej używana metoda wyznaczania wielkości tego marginesu opiera się na klasycznej parametryzacji. W konsekwencji nie są w niej uwzględnione międzyfrakcyjne trendy czasowe, których obecność została wykazana w trakcie statystycznej analizy danych klinicznych. W prezentowanej pracy została zaproponowana nowa formuła obliczania marginesu, uwzględniająca trendy czasowe. W granicy braku tych trendów prezentowana formuła uprasza się do obecnie stosowanej.

Słowa kluczowe:

radioterapia, niepewność ułożenia, dawka skumulowana, protokoły weryfikacji, marginesy CTV-PTV, trendy czasowe, symulacje Monte Carlo

Contents

Acknowledgements

Abstract **I**

Streszczenie **III**

Table of contents **V**

List of Abbreviations **IX**

List of Symbols **XI**

List of Figures **XV**

List of Tables **XXVII**

1 Introduction **1**
1.1 Radiotherapy Process . . . 2
1.2 Imaging during radiotherapy . . . 2
1.3 Individualization . . . 5
1.4 Research hypothesis and aim . . . 5
1.5 The structure of the thesis . . . 6

2 Theoretical Basis **7**
2.1 Systematic and random errors . . . 7
2.2 CTV-PTV margins . . . 13
2.2.1 van Herk formula for CTV-PTV margin calculation . . . 14
2.3 Verification Protocols . . . 17
2.3.1 No Action Level Protocol . . . 17
2.3.2 Extended No Action Level Protocol . . . 18
2.4 Dose Volume Histograms (DVHs) . . . 20
2.4.1 Dose Constraints . . . 20

3 Clinical Data **23**
3.1 Background . . . 23
3.2 Materials . . . 23
3.2.1 Database from Maria Sklodowska-Curie Institute — Oncology Center in Warsaw (COI) . . . 23

3.2.2 Database from Erasmus MC University Medical Center Rotterdam (Erasmus MC) . . . 25
3.3 Methods . . . 25
3.3.1 Testing normality . . . 25
3.3.1.1 Reliability of Shapiro-Wilk test . . . 25
3.3.1.2 Data analysis for single patients . . . 25
3.3.1.3 Data analysis for population of patients . . . 25
3.3.2 Time trends analysis . . . 27
3.3.3 Fitting probability distributions . . . 28
3.4 Results - COI . . . 28
3.4.1 Testing normality . . . 28
3.4.1.1 Data analysis for single patients . . . 28
3.4.1.2 Data analysis for population of patients . . . 28
3.4.2 Time trends analysis . . . 35
3.5 Results - Erasmus MC . . . 43
3.5.1 Testing normality . . . 43
3.5.1.1 Data analysis for single patients . . . 43
3.5.1.2 Data analysis for population of patients . . . 43
3.5.2 Time trends analysis . . . 49
3.5.3 Fitting probability distributions . . . 57
3.5.3.1 Mean Setup Error and SD of setup error . . . 57
3.5.3.2 Time Trend Parameters . . . 57
3.6 Discussion . . . 61

4 Patient population characterization - the choice of right parameters 67
4.1 Background . . . 67
4.2 Materials and methods . . . 67
4.2.1 Synthetic population experiments . . . 67
4.2.2 Parametrization methods . . . 68
4.2.2.1 Conventional Parametrization . . . 68
4.2.2.2 Trendline Parametrization . . . 69
4.2.3 Evaluating quality of parametrization . . . 71
4.2.3.1 Ability of reproducing the initial data . . . 71
4.2.3.2 NAL and eNAL verification protocols performance . . . 71
4.3 Results . . . 73
4.3.1 Synthetic population experiments . . . 73
4.3.1.1 Ability of reproducing the initial data . . . 73
4.3.1.2 NAL and eNAL verification protocols performance . . . 73
4.3.2 Clinical Data . . . 75
4.3.2.1 Ability of reproducing the initial data . . . 75
4.3.2.2 NAL and eNAL verification protocols performance . . . 78
4.4 Discussion . . . 91

5 Estimation of Cumulative Dose 95
5.1 Background . . . 95
5.2 Material and methods . . . 95
5.2.1 Clinical Data . . . 95

5.2.2 Radiotherapy Treatment Plans . . . 95
5.2.2.1 3D-CRT . . . 96
5.2.2.2 IMRT . . . 96
5.2.2.3 VMAT . . . 96
5.2.3 Cumulative Dose Distribution . . . 96
5.2.4 Methods of estimating Cumulative Dose Distribution . . . 97
5.2.4.1 Estimation with mean setup error . . . 97
5.2.4.2 Estimation with mean dose distribution . . . 98
5.2.5 Gamma Evaluation . . . 98
5.2.6 DVHs Evaluation . . . 99
5.2.7 Statistical comparison . . . 99
5.3 Results . . . 100
5.3.1 Gamma Evaluation . . . 100
5.3.1.1 Estimating with mean setup error . . . 100
5.3.1.2 Estimating with mean dose distribution . . . 100
5.3.1.3 Comparison of estimation methods . . . 104
5.3.2 DVHs Evaluation . . . 104
5.3.2.1 Comparison of doses in CTV . . . 104
5.3.2.2 Comparison of doses in PTV . . . 105
5.3.2.3 Comparison of doses in rectum . . . 105
5.3.2.4 Comparison of estimation methods . . . 105
5.4 Discussion . . . 110

6 Proposal of a new off-line verification protocol 111
6.1 Background . . . 111
6.2 Material and methods . . . 112
6.2.1 Individualized No Action Level Protocol (iNAL) . . . 112
6.2.2 Cumulative Average Protocol (CA) . . . 113
6.2.3 Synthetic Data (Monte Carlo patients) . . . 113
6.2.4 Chi-square distribution theoretical background . . . 113
6.2.5 Receiver Operating Characteristic curve . . . 115
6.2.6 Evaluation and comparison of protocols . . . 117
6.3 Results . . . 117
6.3.1 Estimation of average number of images in iNAL protocol . . . 117
6.3.2 Accuracy test (ROC curves analysis) . . . 117
6.3.3 Evaluation and comparison of protocols . . . 126
6.4 Application recipe . . . 134
6.5 Discussion . . . 134

7 CTV-PTV margins in case of time trends 135
7.1 Background . . . 135
7.2 Materials and Methods . . . 135
7.2.1 Margin recipe in case of time trends . . . 135
7.2.1.1 Proof that mean setup error is equal to the middle position . . . 138
7.2.1.2 3D margin calculation . . . 139
7.2.2 Fitting equation for time trend margin calculation . . . 139
7.2.3 Synthetic populations . . . 140

7.2.4 Clinical Data 140
7.2.5 Validation of margin recipe – Ellipsoid test 141
7.3 Results 145
7.3.1 Synthetic populations 145
7.3.2 Clinical Data 145
7.3.3 Fitting equation for time trend margin calculation 145
7.4 Discussion 154
7.5 Look-up tables 156

8 Summary **167**

References **169**

List of Abbreviations

3D-CRT	3D Conformal RadioTherapy
AAA	Analytical Anisotropic Algorithm
AP	Anterior-Posterior
ASCO	American Society for Clinical Oncology
AUC	Area Under ROC Curve
CA	Cumulative Average
CBCT	Cone-Beam CT
CLT	Central Limit Theorem
COI	Maria Sklodowska-Curie Institute — Oncology Center
CT	Computed Tomography
CTV	Clinical Target Volume
DRRs	Digitally Reconstructed Radiographs
DVH	Dose Volume Histogram
EBRT	External Beam RadioTherapy
eNAL	extended No Action Level protocol
EORTC	European Organisation for Research and Treatment of Cancer
Erasmus MC	Erasmus MC University Medical Center Rotterdam
ESTRO	European SocieTy for Radiotheraphy and Oncology
GTV	Gross Tumor Volume
HF	Head-Feet
IMRT	Intensity Modulated Radiation Therapy
iNAL	individualized No Action Level protocol

IV	Irradiated Volume
kV	Kilovolt
LR	Left-Right
MC	Monte Carlo
MLC	Multileaf Collimator
MRI	Magnetic Resonance Imaging
MU	Monitor Units
MV	Megavolt
NAL	No Action Level protocol
NMR	Nuclear Magnetic Resonance
OAR	Organ At Risk
OLS	Ordinary Least Square
OVP	Offline Verification Protocol
PET	Positron Emission Tomography
PTOK	Polish Society for Clinical Oncology (*Polskie Towarzystwo Onkologii Klinicznej*)
PTV	Planning Target Volume
QA	Quality Assurance
RMS	Root Mean Square
RSS	Residual Sum of Squares
ROC	Receiver Operating Characteristic
SAL	Shrinking Action Level
SIB	Simultaneous Integrated Boost
STD	Standard Deviation
TBI	Total Body Irradiation
TPS	Treatment Planning System
TV	Treated Volume
VMAT	Volumetric Arc Therapy

List of Symbols

Symbols written in **bold** represent vectors.

a_p time trend slope for patient p, given in mm/fraction

b_p time trend offset for patient p

CDF Cumulative Distribution Function, CDF($\chi_{STAT}{}^2, dof$) is used to determine the Cumulative Distribution Function for $\chi_{STAT}{}^2$ and dof degrees of freedom

χ_{STAT} statistic of χ^2 test

$\boldsymbol{\Delta_p}$ systematic displacement between planned and treatment isocenter in relative to patient anatomy for patient p during whole treatment

$\boldsymbol{\delta_{p,f}}$ random deviation of displacement between planned and treatment isocenter position in relative to patient anatomy for patient p in fraction f

D_{cum} cumulative dose distribution, i.e. dose distribution which was blurred in order to consider random errors that might appear during each fraction of the treatment

$D_{p,f}(\boldsymbol{m_{p,f}})$ dose distribution in fraction f for patient p calculated with isocenter shifted by $-\boldsymbol{m_{p,f}}$

$D_{p,md}(n)$ Estimation of cumulative dose for patient p done with mean dose distribution of n first fractions

$D_{p,ms}(n)$ Estimation of cumulative dose for patient p done with mean setup error calculated from n first fractions

$D_{threshold}$ dose threshold used for derivation of dose population histograms

DVH_{cum} cumulative dose volume histogram (DVH)

DVH_{diff} differential dose volume histogram (DVH)

$\boldsymbol{E_p}$ systematic measurement error for patient p

$\boldsymbol{\varepsilon_{p,f}}$ random measurement error which can vary from fraction to fraction

F total number of fractions in a treatment course

f one of the treatment fractions given in a treatment course

FN False Negative — falsely predicted negative condition

FP False Positive — falsely predicted positive condition

FPR False Positive Rate

$\gamma(\boldsymbol{r_r})$ gamma parameter used in radiotherapy Quality Assurance (QA). $\gamma(\boldsymbol{r_r})$ depends on accepted dose and distance criteria — ΔD_M and Δd_M, respectively.

$\boldsymbol{m_{p,f}}$ displacement vector measured for patient p in fraction f

$\overline{\boldsymbol{m_p}}$ mean displacement vector of patient p

$\boldsymbol{M_a}$ mean trend line slope for population of patients P

MD Maximum Deviation

mid middle position in a trendline

$\mathbf{M^{PTV}}$ CTV-PTV margin

$\boldsymbol{\mu_{pf}}$ random error of patient p in fraction f

N number of patients p in their population P

P population of patients

p patient, from a population of patients P

PDF Probability Density Function

r superscript used to denote random errors component, e.g. m_p^r, a_p^r

r_f random error in fraction f (residual error)

$\boldsymbol{S_p}$ systematic error for patient p, if written without vector it refers to one of principal axis. Note that $S_p = \overline{\boldsymbol{m_p}}$

SD_{SD} width of distribution of SD_p other population of patients P

SEM standard error of mean

$SEM_{\overline{m_p}}(n)$ standard error of mean displacement for patient p calculated from n fractions

$\boldsymbol{\Sigma_a}$ standard deviation of mean trend line slope for population of patients P

$\boldsymbol{SD_p}$ standard deviation of mean setup error of patient p

$\boldsymbol{\Sigma_d}^2$ variance of the delineation error

$\boldsymbol{\sigma_m}^2$ variance of the organ motion during treatment execution

$\boldsymbol{\Sigma_m}^2$ variance of the organ motion during treatment preparation

$\boldsymbol{\sigma_p}$ beam penumbra width

$\boldsymbol{\sigma_s}^2$	variance of the setup error during treatment execution
$\boldsymbol{\Sigma_s}^2$	variance of the setup error during treatment preparation
$\boldsymbol{\Sigma}$	variation of systematic errors calculated for each patient p in population P
$\boldsymbol{\sigma}$	Root Mean Square of SD_p
$\boldsymbol{\Sigma_{res}}$	variation of residual systematic error for whole patient population
TNR	True Negative Rate
TN	True Negative — truly predicted negative condition
TPR	True Positive Rate
TP	True Positive — truly predicted positive condition
tt	superscript used to denote time trend residual errors, i.e. errors defined through their distance from the trendline
V_{total}	total structure volume
x_p	trendline error, i.e. non-random error in case of time trend

List of Figures

1.1 Photography of the MV imaging system in use. The radiation beam goes from the linear accelerator itself (accelerator head is seen on the left), crosses the patient (here anthropomorphic phantom) and is detected by a flat panel (on the right). Attenuation coefficients over the whole beam path are measured. Photography was taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 3

1.2 Photography of the kV imaging system in use. The radiation beam goes from the X-ray tube (on the left), crosses the patient (here anthropomorphic phantom) and is detected by a flat panel (on the right). Attenuation coefficients over the whole beam path are measured. Photography was taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 3

1.3 Comparison of kV (on left) and MV (on right) imaging done for anthropomorphic phantom (see Fig. 1.1-1.2). Upper images show anterior-posterior imaging, lower images show left-right imaging. Scans were taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 4

1.4 Comparison between planning CT and CBCT. Structures seen on planning CT are: yellow – CTV, red – PTV, green – rectum. Scans were taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 5

2.1 Photography presents a patient lying on a treatment couch in the individualised thermoplastic mask (yellow) for head and neck region immobilization. Laser system (red) crossing on adhesive tapes attached to the mask can be also seen. Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 8

2.2 Photography presents a patient lying on a treatment couch in the individualised vacuum bag (blue) with additional knee support (green). Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 8

2.3 Laser system (red) in pelvic region. Marker can be seen as a small black dot on the laser crossing. Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 9

2.4 Methodology of transformation of laser crossing point into CT. Radio-opaque markers are placed on patient body prior to CT scan (left figure) in order to represent tattoos. These markers can be seen on CT exam slice (right figure). Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 10

2.5 Influence of setup errors on dose distribution. The black circle indicates the target volume. The dot indicates the planned isocenter position. Vector shown on 2.5b and 2.5d indicates the systematic displacement of the treatment isocenter relative to the planned one. 10

2.6 Setup images performed prior to treatment aligned to DRRs on which bony structures are delineated. Images were acquired in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 11

2.7 Schematic illustration of volumes used in prescribing and reporting radiotherapy (graph derived from my Master Thesis [32]). 13

2.8 Figure presents the methodology of creating probability histogram of the cumulative dose over a population of patients (dose-population histogram). For simplicity CTV consisting of one point is assumed. Fig. 2.8a shows how dose-population histogram is obtained. Red curve on upper graph presents dose in point CTV in relation to error in CTV position. We have to decide on $D_{threshold}$ which should be delivered (e.g. the minimum dose in CTV, here 95%). Then we look into the range/collection of all possible CTV positions in which the dose would be at least $D_{threshold}$. Knowing this collection we check what fraction of whole population it is (in other words we check the probability that CTV would have this location) — see lower graph of Fig. 2.8a for probability of CTV displacement (CTV position error). That procedure leads us to a single point in dose-population histogram (Fig. 2.8b) and therefore has to be repeated for each $D_{threshold} \epsilon [0\%; 100\%]$. Dose-population histogram prepared for data shown here is presented in Fig. 2.8b. Selected point (95%, 90%) shows that 95% of dose would be achieved in 90% of patients. These figure was created on the basis of van Herk paper [90]. 16

2.9 Example of NAL protocol performance. Figure shows NAL in case when mean setup error (i.e. the correction value) is well estimated on the basis of 3 first fractions. Arrow shows an example shift between corrected an uncorrected Setup Error. 18

2.10 Example of eNAL protocol performance in presence of inter-fraction time trend. First three fractions (black dots) are irradiated without any correction. Afterwards a correction is calculated and applied for five upcoming fractions (first arrow indicates this correction value). In 8th fraction an imaging is performed and a linear fit is done for uncorrected setup errors. On the basis of this fit correction the value for upcoming five fractions is recalculated. The imaging is done in each fifth fraction, followed by a linear fit and update of correction factor. Arrows show first fractions after the new correction value was calculated and dotted black lines show the correction value and fractions in which it was applied. It can be seen how the correction value is changing in time. 19

2.11 Simple 2D dose matrix with a PTV contour (red circle) is presented. While considering $D_i = 40$ Gy, for a differential histogram only two voxels (marked in light red) would be taken into account. For a cumulative histogram, while considering $D_i = 40$ Gy, all voxels with $D \geq 40$Gy (marked in light red) would be counted. 21

2.12 Explanation on how differential DVH is created. The classical histogram is created from the dose matrix at first (2.12a). Afterwards the middle positions of bars are taken and joined with line (2.12b). 21

2.13 Cumulative Dose Volume Histogram. For some dose-volume constraints cubic centimeters are used, more often volume is given in percentage of total structure volume. 22

2.14 Illustration of dosimetric parameters. In order to check $D_V = D_{95\%}$ we look on volume $V = 95\%$ on the y-axis, search for point on a curve and check adequate dose D (Fig. 2.14a). In order to check $V_D = V_{30\ \mathrm{Gy}}$ we look on dose $D = 30$ Gy on the x-axis, search for point on a curve and check adequate volume V (Fig. 2.14b). 22

3.1 CT scan for prostate cancer patient. Knee support is shown. Photography was taken with patient permission in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw. 24

3.2 Histogram of p-values for setup error in left-right direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation. 26

3.3 Histogram of p-values for setup error in anterior-posterior direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation. 26

3.4 Histogram of p-values for setup error in head-feet direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation. 27

3.5 Histogram of setup errors distributions in left-right direction (3.5a and 3.5b), anterior-posterior direction (3.5c and 3.5d) and head-feet direction (3.5e and 3.5f). In the left column distribution of setup errors with the lowest p-value is presented. In the right column distribution of setup errors with the highest p-value is presented. Data achieved for COI database. 29

3.6 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in left-right direction. Data achieved for COI database. 30

3.7 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in anterior-posterior direction. Data achieved for COI database. 31

3.8 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in head-feet direction. Data achieved for COI database. 31

3.9 Comparison of histograms of p-values achieved with Shapiro-Wilk test for setup errors in all directions. Data achieved for COI database. 32

3.10 Histogram of minimum p-value, out of three directions, for all patients. Data achieved for COI database. 32

3.11 Number of patients with probable not normal setup error distribution in 0-1-2-3 directions. Confidence interval was set to 95%. Data achieved for COI database. 33

3.12 Histograms of $\overline{m_p}$ and SD_p for setup errors in left-right direction (3.12a and 3.12b), anterior-posterior direction (3.12c and 3.12d) and head-feet direction (3.12e and 3.12f). The p-value of Shapiro-Wilk test is written on each histogram. Data achieved for COI database. 34

3.13 Plots of setup error distribution along left-right axis (3.13a and 3.13b), anterior-posterior axis (3.13c and 3.13d) and head-feet axis (3.13e and 3.13f). On the left patients with the lowest observed absolute slope values are shown (i.e. patients with no observed time trend), on the right patients with highest observed slope values are shown (significant time trend observed). LR_{se}, AP_{se}, HF_{se} stands for m_{pf} in particular direction while n_f describes fraction number and is a measure of time. Data achieved for COI database. 36
3.14 Histograms of slope (left) and offset (right) values for time trend OLS analysis of setup errors in each direction. Data achieved for COI database. 37
3.15 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database. 38
3.16 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database. 38
3.17 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database. 39
3.18 Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in left-right direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database. 39
3.19 Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in anterior-posterior direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database. 40
3.20 Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in head-feet direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database. 40
3.21 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Left-Right direction. Data achieved for COI database. 41
3.22 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Anterior-Posterior direction. Data achieved for COI database. 41
3.23 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Head-Feet direction. Data achieved for COI database. 42
3.24 Histograms of number of patients with specified number of directions with probable not normal distribution — comparison between pure setup errors and time trend residuals. Data achieved for COI database. 42
3.25 Histograms of setup error distribution in left-right direction (3.25a and 3.25b), anterior-posterior direction (3.25c and 3.25d) and head-feet direction (3.25e and 3.25f). Data achieved for Erasmus MC database. 44
3.26 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in left-right direction. Data achieved for Erasmus MC database. 45
3.27 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in anterior-posterior direction. Data achieved for Erasmus MC database. 45
3.28 Histogram of p-values achieved with Shapiro-Wilk test for setup errors in head-feet direction. Data achieved for Erasmus MC database. 46

3.29 Comparison of histograms of p-values achieved with Shapiro-Wilk test for setup errors in all directions. Data achieved for Erasmus MC database. 46
3.30 Histogram of minimum p value, out of three directions, for all patients. Data achieved for Erasmus MC database. 47
3.31 Number of patients with probable not normal setup error distribution in 0-1-2-3 directions. Confidence interval was set to 95%. Data achieved for Erasmus MC database. 47
3.32 Histograms of $\overline{m_p}$ and SD_p for setup errors in left-right direction (3.32a and 3.32b), anterior-posterior direction (3.32c and 3.32d) and head-feet direction (3.32e and 3.32f). The p-value of Shapiro-Wilk test comparing normal ditribution with presented histogram is written on each figure. Data achieved for Erasmus MC database. 48
3.33 Plots of setup error distribution along left-right axis (3.33a and 3.33b), anterior-posterior axis (3.33c and 3.33d) and head-feet axis (3.33e and 3.33f). On the left patients with the lowest observed absolute slope values are shown (with no observed time trend), on the right patients with highest observed slope values are shown (significant time trend observed). LR_{se}, AP_{se}, HF_{se} stands for m_{pf} in particular direction while n_f describes fraction number and is a measure of time. Data achieved for Erasmus MC database. 50
3.34 Histogram of slope (left) and offset (right) values for time trend OLS analysis of setup errors in each direction. Data achieved for Erasmus MC database. 51
3.35 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database. 52
3.36 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database. 52
3.37 Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database. 53
3.38 Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in left-right direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database. 53
3.39 Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in anterior-posterior direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database. 54
3.40 Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in head-feet direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database. 54
3.41 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in left-right direction. Data achieved for Erasmus MC database. 55
3.42 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in anterior-posterior direction. Data achieved for Erasmus MC database. 55

3.43 Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in head-feet direction. Data achieved for Erasmus MC database. 56
3.44 Comparison of number of patients with non-normal distribution of pure/time trend residual setup errors in 0-1-2-3 directions. Confidence level was set to 95%. Data achieved for Erasmus MC database. 56
3.45 Comparison between histograms of mean setup errors and SD of setup errors and fitted distributions. Data achieved for Erasmus MC database. 58
3.46 Comparison between trendline slope/offset histograms and fitted distributions. Data achieved for Erasmus MC database. 59
3.47 Comparison between histograms of SD of Time Trend residual setup errors and fitted distributions. Data achieved for Erasmus MC database. 60
3.48 SD of slope changing with number of fractions used in OLS trend fit. Histogram of achieved slope values for 39 fractions and $SD = 3$ mm is also shown. 62
3.49 Comparison of slope distribution achieved for Erasmus MC database in left-right direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends. 63
3.50 Comparison of slope distribution achieved for Erasmus MC database in anterior-posterior direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends. 64
3.51 Comparison of slope distribution achieved for Erasmus MC database in head-feet direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends. 64
3.52 Histogram of observed total trend motion (i.e. change in setup errors related only to trendline through the entire treatment). Data achieved for Erasmus MC database. 65
3.53 Comparison of two CT slices of the same patient. On the left the situation in which muscles' tension can be observed. On the right — relaxed muscles. The change in the PTV position (red line) can be observed — PTV is higher in regard to table top while muscles are tighten. GTV is shown as a yellow region while rectum is a green one. 65

4.1 Histogram of left-right mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database. 70
4.2 Histogram of anterior-posterior mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database. 70
4.3 Histogram of head-feat mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database. 71
4.4 Correlation between slope (left) and offset (right) for time trend OLS analysis and mean value of setup errors in each direction. Data achieved for Erasmus MC database. 72
4.5 Comparison of Σ_{res} achieved with NAL protocol for five arbitrary chosen synthetic and MC populations. Direct simulation shows the NAL performance for synthetic population, it is ground truth for this evaluation. 74

4.6 Values of Σ_{res} achieved for 36 different synthetic populations (direct simulation) and adequate MC populations: with standard parametrization (Conventional MC) and trendline parametrization (Trendline MC). In each case NAL protocol was applied after first 3 fractions with imaging. Total treatment time was set to 39 fractions. 76
4.7 Values of Σ_{res} achieved for 36 different synthetic populations (direct simulation) and adequate MC populations: with standard parametrization (Conventional MC) and trendline parametrization (Trendline MC). In each case eNAL protocol was applied in which first correction was done after third fraction with update of correction factor done weekly. Total treatment time was set to 39 fractions. . 77
4.8 Comparison between histograms of mean setup error achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 79
4.9 Comparison between histograms of setup errors SD achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 80
4.10 Comparison between histograms of d_{100} achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 81
4.11 Comparison between histograms of time trend slope achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 82
4.12 Comparison between histograms of time trend offset achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 83
4.13 Comparison between histograms of mean setup error achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 84
4.14 Comparison between histograms of setup errors SD achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 85
4.15 Comparison between histograms of d_{100} achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 86
4.16 Comparison between histograms of time trend slope achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 87

4.17 Comparison between histograms of time trend offset achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed. 88

4.18 Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in left-right direction. 89

4.19 Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in head-feet direction. 90

4.20 Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in anterior-posterior direction. 90

4.21 Comparison of Σ_{res} achieved with NAL protocol for arbitrary chosen synthetic populations (direct simulation). Performance in case of inverted fraction order is also shown 92

4.22 Examination of $\sigma/0.5\sqrt{\pi}$ changes in time. 93

5.1 Treatment plan geometry for 3D-CRT (Fig. 5.1a) and IMRT (Fig. 5.1b) plans. . 96

5.2 Distribution of isocenter positions for 3D-CRT (Fig. 5.2a) and IMRT (Fig. 5.2b) plans. 97

5.3 Schematic illustration of the concept of gamma evaluation method. For readability distance space is set to 2D. Point $(\boldsymbol{r_r}, D_r)$ stands for point in reference dose distribution. Point $(\boldsymbol{r_c}, D_c)$ stands for exemplary point from compared dose distribution. ΔD_M denotes dose difference tolerance, while Δd_M denotes maximum allowed distance to agreement (graph adapted from Depuydt et al. [19]). 99

5.4 Examples of gamma evaluation result for 3D-CRT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient. 101

5.5 Mean percent of voxels with gamma index not greater than 1 ($\gamma \leqslant 1$) for comparison between $D_{p,cum}$ and $D_{p,ms}(n)$. Standard deviation is presented as error. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right). 101

5.6 Examples of gamma evaluation result for IMRT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient. 102

5.7 Mean percent of voxels with gamma index not greater than 1 ($\gamma \leqslant 1$) for comparison between $D_{p,cum}$ and $D_{p,md}(n)$. Standard deviation is presented as error. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right). 102

5.8 Examples of gamma evaluation result for VMAT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient. 103
5.9 Number of patients with gamma index not greater than 1 ($\gamma \leqslant 1$) in less than 95% (90%) of voxels for global (local) gamma evaluation. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right). 103
5.10 Results of Mann-Whitney test performed for different n fractions used in dose estimation. Gamma index passing rate for $D_{p,ms}(n)$ and $D_{p,md}(n)$ were compared. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right).104
5.11 Comparison of profiles for $D_{p,cum}$, $D_{p,ms}(n)$ and $D_{p,md}(n)$. It can be seen that $D_{p,ms}(n)$ has the steepest dose gradient in penumbra region. 105
5.12 Comparison of dosimetric parameters for CTV between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in D_{min}, D_{mean} and $D_{98\%}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}. 106
5.13 Comparison of dosimetric parameters for PTV between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in D_{min}, D_{mean} and $D_{98\%}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}. 107
5.14 Comparison of dosimetric parameters for rectum between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in $V_{60\text{ Gy}}$ and $V_{65\text{ Gy}}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}. 108
5.15 Mann-Whitney sign rank test results for dose parameters in CTV and PTV. . . 109
5.16 Mann-Whitney sign rank test results for dose parameters in rectum. 110

6.1 Example of of wrong estimation of mean setup error in NAL protocol. Arrow shows an example shift between corrected an uncorrected setup error. 112
6.2 Probability density functions of χ^2 distribution with different degrees of freedom: 3 ($n_f = 4$) — left and 4 ($n_f = 5$) — right. The shaded area shows the probability of observing $SEM_{\bar{m}_p}$ equal or bigger than the flag. 115
6.3 Probability density function of χ^2 distribution with 4 degrees of freedom ($n_f = 5$). The whole shaded area shows the probability of observing $SEM_{\overline{m}_p}$ equal or bigger than the flag if the $SEM_o(n_f) = \frac{{\sigma_o}^2}{(n_f-1)}$. The blue area shows the probability of observing $SEM_{\overline{m}_p}$ equal or bigger than the flag if the $SEM_o(n_f) = \frac{{\sigma_o}^2}{n_f}$. Cyan area shows the probability of having exactly n_f fractions with images. 116
6.4 Histograms of number of patients with specified number of images. Theoretical prediction of χ^2 model is shown with cyan circles. Calculations done for 25 fractions treatment. 118
6.5 Histograms of number of patients with specified number of images. Theoretical prediction of χ^2 model is shown with cyan circles. Calculations done for 39 fractions treatment. 119

6.6 Distributions of number of images for population of patients with constant SD_p. Theoretical prediction of average number of images is shown with yellow diamond. Achieved mean value is shown with red square. Calculations done for 25 fractions treatment. 120
6.7 Distributions of number of images for population of patients with constant SD_p. Theoretical prediction of average number of images is shown with yellow diamond. Achieved mean value is shown with red square. Calculations done for 39 fractions treatment. 121
6.8 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Width of Gaussian distribution of SD_p was taken as $SD_{SD} = 0.26 \cdot \sigma$. Calculations done for 25 fractions treatment. 122
6.9 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Width of Gaussian distribution of SD_p was taken as $SD_{SD} = 0.26 \cdot \sigma$. Calculations done for 39 fractions treatment. 122
6.10 The ROC curves achieved for $\Sigma = 2$ mm, $\sigma = 2$ mm and $SD_{SD} = 1.0$ mm. Left image shows ROC where the classifier was $SEM_{\overline{m_p}}(n_f = 3)$. Right image shows ROC curves for different classifiers: $SEM_{\overline{m_p}}(n_f)$, arrow indicates the direction of increasing n_f. 123
6.11 AUC dependence on number of fractions with imaging — n_f, for $\Sigma = 2$ mm. Dependencies on σ (left figure) and SD_{SD} (right figure) are presented. 123
6.12 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment. 127
6.13 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment. 127
6.14 Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: CA and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 25 fractions treatment. 128
6.15 Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: CA and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 39 fractions treatment. 129
6.16 Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: eNAL and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 25 fractions treatment. 130

6.17 Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: eNAL and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 39 fractions treatment. 131

6.18 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment. 132

6.19 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment. 132

6.20 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment. 133

6.21 Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment. 133

7.1 Example of setup errors for one patient. Trendline is shown. Red arrow indicates one of residual errors (i.e. errors due to trend line). Black arrow indicates one of the trendline errors (i.e. difference between zero position and trendline). Illustration of the concept of maximum deviation is also provided: green arrow shows middle position value, while grey one shows trend change for half of treatment. 136

7.2 Distribution of percentage of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients. 142

7.3 Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients. 142

7.4 Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients. 143

7.5 Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients. 143

7.6 Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients. 144

7.7 Comparison of percentage of patients with at least one trendline error lying outside the margin for van Herk and the proposed margin recipe. Results achieved for isotropic populations with different time trend simulation parameters ($M = 0$ mm, Σ, M_a, Σ_a and σ'). 146
7.8 Percentage of patients with at least one trend setup error lying outside CTV-PTV 3D margin for two compared margin recipes: van Herk (x axis) and the time trend margin (y axis). In total 290 populations were analyzed. 147
7.9 Comparison between margin calculation methods (Python code vs direct equation) for 80% of patients assumed to have all non-random error within the margin. 149
7.10 Comparison between margin calculation methods (Python code vs direct equation) for 85% of patients assumed to have all non-random error within the margin. 149
7.11 Comparison between margin calculation methods (Python code vs direct equation) for 90% of patients assumed to have all non-random error within the margin. 149
7.12 Comparison between margin calculation methods (Python code vs direct equation) for 95% of patients assumed to have all non-random error within the margin. 150
7.13 Comparison between margin calculation methods (Python code vs direct equation) for 99% of patients assumed to have all non-random error within the margin. 150
7.14 Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin. 151
7.15 Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin. 152
7.16 Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin. $\Sigma_m = \Sigma$ 153
7.17 Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in left-right direction. 155
7.18 Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in head-feet direction. 155
7.19 Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in anterior-posterior direction. 156

List of Tables

3.1 Parameters achieved for Erasmus MC database used to check whether observed trends might have a physiological nature. 65

4.1 Table presents p-values of Kolmogorov-Smirnov for comparison of clinical database distributions with different methods of parametrizing and simulating setup errors. In brackets distribution used in MC simulation for selection of SD_p is given. p-values ≥ 0.001 are bold and indicate for which parameter distribution achieved with tested parametrization was comparable to clinical data (i.e. null hypothesis could not been rejected). 75

4.2 Table presents Σ_{res} values achieved for Erasmus MC database in left-right (LR), anterior-posterior (AP) and head-feet (HF) direction after the NAL ($N_m = 3$) and eNAL protocols were simulated. All presented values are given in mm. . . . 89

6.1 Values of AUC for $n_f = 3$, $\Sigma = 1$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm. 123

6.2 Values of AUC for $n_f = 3$, $\Sigma = 2$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm. 124

6.3 Values of AUC for $n_f = 3$. $\Sigma = 3$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm. 124

6.4 Values of AUC for $n_f = 3$, $\Sigma = 1$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm. 124

6.5 Values of AUC for $n_f = 3$, $\Sigma = 2$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm. 125

6.6 Values of AUC for $n_f = 3$. $\Sigma = 3$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm. 125

7.1 Comparison of calculated parameters for standard and time trend parametrization. Mean position of standard parametrization is the same as middle position of time trend parametrization. 139

7.2 Validation of time trend margin recipe for different percentile values used to margin size definition. For each percentile more than 13310 combinations of time trend parameters (Σ, Σ_a) were used. For every combination of parameters 10^5 patients with their setup errors were simulated. First row gives mean achieved value and its standard deviation. Second row provides range of achieved values. 141

7.3 Comparison between van Herk's and the proposed margin recipes in terms of margin size and percentage of patients with at least one trendline error lying outside the 3D margin. Values were calculated for real patient data from Erasmus MC database. Percentage of patients outside the margin was calculated with Ellipsoid Test. 147

7.4 Fit parameters for margin equation (Eq. (7.12)) as a function of percentile of patients with all non-random errors inside the CTV-PTV margin. The Δmargin expresses mean difference of comparison between margin calculation with the Python code and direct calculation using Eq. (7.12), additionally the range is given in brackets. For comparison purposes the last column contains α values as derived by van Herk et al.[90]. 148

7.5 Margin values for requirement of 80% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 80% gives $2.16\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction. 157

7.6 Margin values for requirement of 85% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 85% gives $2.31\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction. 157

7.7 Margin values for requirement of 90% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 90% gives $2.50\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction. 158

7.8 Margin values for requirement of 95% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 95% gives $2.79\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction. 158

7.9 Margin values for requirement of 99% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 99% gives $3.36\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction. 159

7.10 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = 0$ mm. Values of Σ_m and Σ_A are given in mm. 160

7.11 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 0.5$ mm. Values of Σ_m and Σ_A are given in mm. 161

7.12 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 1.0$ mm. Values of Σ_m and Σ_A are given in mm. 162
7.13 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 1.5$ mm. Values of Σ_m and Σ_A are given in mm. 163
7.14 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 2.0$ mm. Values of Σ_m and Σ_A are given in mm. 164
7.15 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 2.5$ mm. Values of Σ_m and Σ_A are given in mm. 165
7.16 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 3.0$ mm. Values of Σ_m and Σ_A are given in mm. 166

Chapter 1

Introduction

Question about life, its borders, and sense of suffering is present in human life since thousands of years, though nowadays that discussion seems to be a bit different. It is hard to argue that a particular person would like to live as long as possible while being in good and healthy condition [22]. Most of us know someone in their family or at least in close relations who got cancer. As cancer is one of most frequent reasons of death [45], being diagnosed with a cancer is often taken as a sentence. That is why so much effort is taken to find out the best cure. Many different types of treatment, such as surgery, chemotherapy, radiotherapy, hormonotherapy and immunotherapy are possible.

Surgery may seem to be the best option as one gets rid of the cancer at all, but it cannot be always performed (e.g. due to the cancer site and/or patient general health state). Moreover, sometimes surgery ends up with a positive margin (carcinogenic cells are found in the material removed from body). That is why in many cases after surgery further additional treatment is required.

Chemotherapy bases on a strong drugs treatment which can stop cell divisions, slow down their growth and lead to death of cancer cells diffused other the body. Therefore it seems to be a good method, especially for patients with higher cancer stage and metastases in lymph nodes. Chemotherapy influences also normal cells. That is why it often leads to stop hair growth.

Radiotherapy uses radiation in order to treat cancer. The source of radiation can be inside the patient (brachytherapy) or outside the patient (External Beam RadioTherapy (EBRT) also called teletherapy). Radiation can cause DNA damage and therefore may lead to cell death. It is impossible to irradiate only the target volume, so there is always a probability of side effects. Usually the radiotherapy is performed for well localized cancers. However, there are procedures like Total Body Irradiation (TBI) in which the whole body is irradiated.

Hormonotherapy is based on the fact that growth and extent on many cancers is related to hormones existence. That is why cancer can sometimes be cured or, more often, stopped or slowed down by hormones. Hormonotherapy is often a part of combined treatment.

The aim of immunotherapy is to modulate the immune system to force it to fight cancer. One of the ideas is to identify antigens present on cancer cells and create antibody for that particular antigen.

The choice of treatment, in general, depends on the stage and type of cancer, its place in body and patient general health. There are many protocols on how to treat each cancer site. One can easily find recommendations given for example by Polish Society of Clinical Oncology (PTOK – *Polskie Towarzystwo Onkologii Klinicznej*) [44], American Society for Clinical Oncology (ASCO) [69] or European Organisation for Research and Treatment of Cancer (EORTC)

[26]. A combined treatment is often performed. Decision on treatment type should be made by a multidisciplinary group. The patient preference should also be taken into consideration.

Each type of treatment mentioned above should be applied in a manner as precise as possible. Still, there may appear some errors which may influence the outcome of the treatment. Therefore they should be taken into account and minimized.

In the thesis presented here I studied patient setup error during EBRT and its influence on the doses received during irradiation. As I will concentrate only on EBRT, in the remainder of the Thesis the "radiotherapy" term would be used to describe it.

1.1 Radiotherapy Process

The aim of radiotherapy is to deliver high, usually uniform dose to the target volume while sparing surrounding healthy tissues. In order to do that one has to choose patient position on treatment couch — for example whether the patient should lie in prone or supine position. During the patient preparation it should be decided if any support, like knee wedge, vacuum bag or mask, should be used [10, 50, 64, 66, 68, 77, 81, 86]. After patient position is established the Computed Tomography (CT) scan is performed. In the case of cancer localized in pelvis region protocols of bladder fulfillment or rectum purification are often applied prior to CT scan. After CT scans are performed, physicians and/or radiographers have to delineate target volumes and organs at risk (OARs). The CT scans prepared in such a way are then used in Treatment Planning System (TPS) in order to choose treatment parameters and calculate dose distribution. After the treatment plan is completed and accepted by a physician the irradiation may start.

1.2 Imaging during radiotherapy

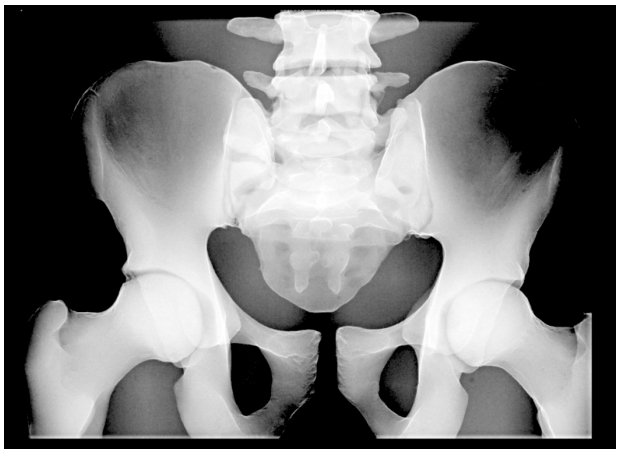
During radiotherapy we have to deal with many sources of uncertainties and errors. Much effort is taken to minimize the uncertainty of not delivering the right amount of dose into the right volume. One of the most widely applied methods is to check the patient position (setup) which should be as similar as possible to the position during the CT scan used for planning radiotherapy. That is why modern medical accelerators (often called linacs) are equipped with imaging systems which can be used to check and monitor patient position. The easiest way is to use megavolt (MV) beam for patient positioning (see Fig. 1.1). In that case linac should have the flat panel detector for MV imaging. Another solution is to use an additional classical X-ray tube lamp, this is called the kilovolt (kV) imaging and also requires dedicated flat panel (see Fig. 1.2). As the energy used in kV imaging is much lower than in MV imaging the contrast of obtained setup images is better (see Fig. 1.3).

Despite the fact which modality (kV or MV) is used, the image have to be compared with a reference image in order to define the patient setup error. Usually these reference image would be generated from the CT and therefore would be called Digitally Reconstructed Radiograph (DRR).

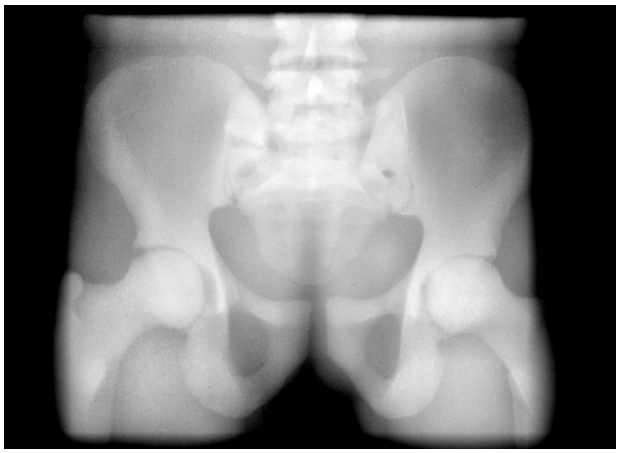
Additionally some linacs have a possibility of acquiring the so called Cone-Beam CT (CBCT) which is claimed to be a better choice at least for some treatment sites [56]. In that case a 3D volume comparison is done between CBCT and planning CT (see Fig. 1.4b).

Figure 1.1: Photography of the MV imaging system in use. The radiation beam goes from the linear accelerator itself (accelerator head is seen on the left), crosses the patient (here anthropomorphic phantom) and is detected by a flat panel (on the right). Attenuation coefficients over the whole beam path are measured. Photography was taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

Figure 1.2: Photography of the kV imaging system in use. The radiation beam goes from the X-ray tube (on the left), crosses the patient (here anthropomorphic phantom) and is detected by a flat panel (on the right). Attenuation coefficients over the whole beam path are measured. Photography was taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

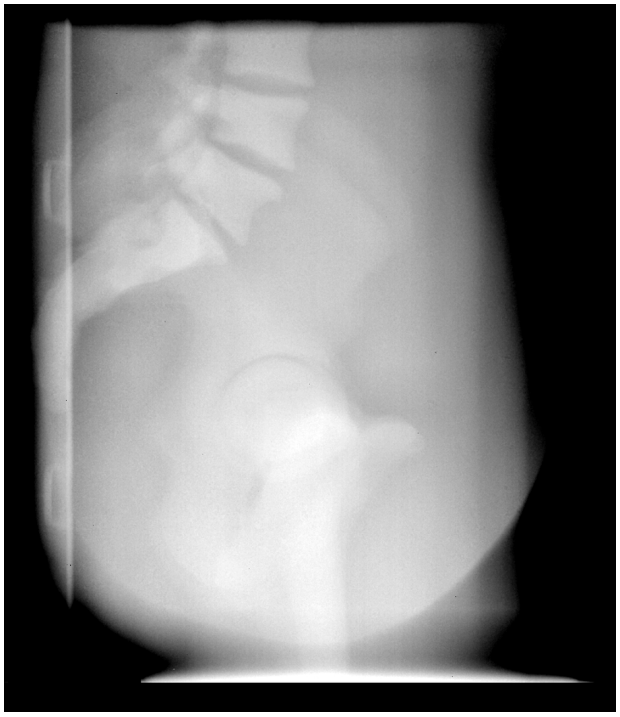


(a) Anterior kV image



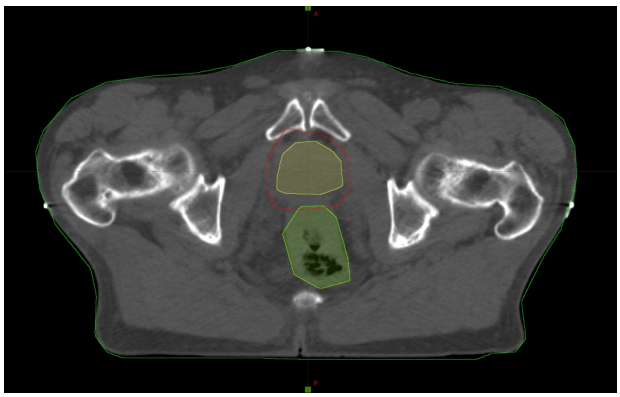
(b) Anterior MV image

(c) Lateral kV image

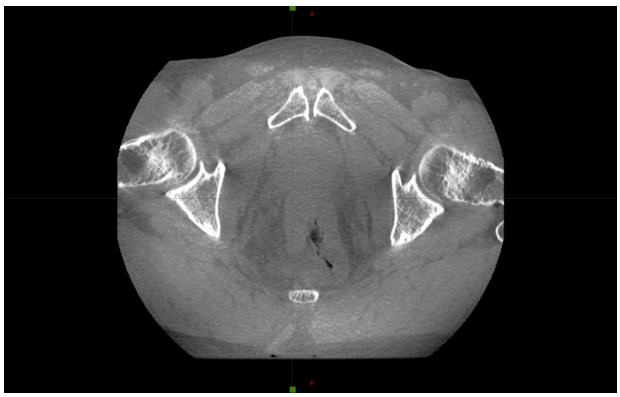


(d) Lateral MV image

Figure 1.3: Comparison of kV (on left) and MV (on right) imaging done for anthropomorphic phantom (see Fig. 1.1-1.2). Upper images show anterior-posterior imaging, lower images show left-right imaging. Scans were taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.



(a) Planning CT



(b) CBCT aquired prior to treatment

Figure 1.4: Comparison between planning CT and CBCT. Structures seen on planning CT are: yellow – CTV, red – PTV, green – rectum. Scans were taken in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

1.3 Individualization

European SocieTy of Radiotherapy and Oncology (ESTRO) wrote in its vision 2020: "Every cancer patient in Europe will have access to state-of-the-art radiation therapy as part of a multidisciplinary approach where treatment is individualized for the specific patients cancer, taking account of the patient's personal circumstances." These circumstances are related among other things to prescribed dose, dose fractionation and dose distribution [46]. Due to the fact that each patient has different setup errors we can think of individualization of treatment execution. One of the approaches is Adaptive RadioTherapy (ART) [24, 96]. Another approach is plan of the day [31, 41]. Also new treatment planning concepts are investigated like probability planning or margin-less planning [5, 27, 28]. The aim of individualization during radiotherapy process is to decrease volume of normal tissues receiving high dose with the dose in target remaining the same.

1.4 Research hypothesis and aim

The influence of patients setup errors on the doses received by them during radiotherapy was widely analyzed by different groups [4, 7, 11, 12, 15, 90, 89]. These studies concerned cumulative dose distribution, verification protocols and CTV-PTV margin concepts. In most of these studies it was assumed that setup errors have Gaussian distribution. In the presented thesis that assumption was checked in two different sets of clinical data. Influence of these findings on previously mentioned concepts was investigated. Propositions were made how to overcome the issues found.

As mentioned in Section 1.3 there is much effort done nowadays in order to individualize radiotherapy treatment. These methods often require much workload — like imaging done everyday with on-line registration. This requires additional time on a machine and people resources to perform such procedures. It is also important to mention that with longer session time the higher is the chance of the patient movement during the session.

Not all hospitals can afford additional workload to provide an individualized treatment. Also new treatment planning algorithms are often still not clinically implemented or require

additional budget for their license.

Last but not least, some patients have two or more targets irradiated at the same time. These targets do not always behave (in terms of motion during radiotherapy) in the same manner [29, 72, 93]. In such a case on-line verification procedures can fit only one of the targets.

The hypothesis standing prior to the performed studies was that it is possible to propose simple forms of radiotherapy individualization not requiring much increase in workload. In order to check this hypothesis an analysis of large clinical data was performed. Afterwards the currently applied methods and concepts (like setup errors parametrization, verification protocols, margins) were investigated and alternative solutions were proposed. Aim of this alternative solutions was to overcome obstacles which were seen during analysis and provide as simple and individualized approaches as possible not requiring much additional workload.

1.5 The structure of the thesis

At the beginning of this thesis I will start with providing some necessary theoretical basis – see Chapter 2. In Chapter 3 I will statistically analyze two populations of prostate cancer patients in terms of their setup errors. First group would consist of 100 patients treated in the Maria Sklodowska-Curie Institute — Oncology Center. Positioning of that group of patients was done on the basis of bony anatomy. Second group of patients would consist of 835 patients treated in Erasmus MC University Medical Center Rotterdam (Erasmus MC) in the Netherlands. Positioning of that group of patients was done on the basis of gold markers. On the basis of statistical analysis of clinical data I will discuss methods of Monte Carlo (MC) Simulation of population of patients – see Chapter 4. In Chapter 5 I will discuss methods of calculating and estimating the cumulative dose distribution for a particular patient. Next I will propose a new off-line verification protocol (OVP) and compare it with existing ones, previously described in Chapter 2. Theoretical basis of the new OVP would be discussed and MC simulations results would be provided – see Chapter 6. In Chapter 7 I will discuss problems while using CTV-PTV margin calculated with van Herk recipe [90] in presence of time trends. I will propose some new solutions; a validation and comparison with van Herk recipe would be also provided. The Thesis would be finally concluded in Chapter 8.

Chapter 2

Theoretical Basis

2.1 Systematic and random errors

Radiotherapy, like every other complex process may be divided into certain steps. During each step errors and uncertainties may occur. Sources of that errors were widely discussed in literature as well as their systematic and random nature [1, 4, 58, 89].

In order to understand sources of errors during radiotherapy one has to keep in mind all steps it consists of. After the decision for radiotherapy is made, patient's position during treatment has to be chosen (e.g. anterior/posterior). If needed, additional support equipment, like thermoplastic masks (see Fig. 2.1) or vacuum bag (see Fig. 2.2), should be prepared.

Following decision on patient's position a Computed Tomography (CT) scan has to be performed. A laser system is used to made patient's position reproducible. Each CT and therapeutic room is equipped with two lateral lasers (vertical and horizontal one placed on walls parallel to couch) and one sagittal laser (placed on wall opposite to CT/treatment machine). The points on the patient's body, where two lasers cross, have to be marked. In the pelvis and chest region the point tattoos on the skin are usually done. In the head and neck and brain region laser positions are drawn on adhesive tapes put on the patient's mask. (see Fig. 2.1 and 2.3).

As one has to know where the laser crossings points are on the patient's CT, radio-opaque markers have to be put on the patient's body before the scan is performed (see Fig. 2.4). The CT scan is the first place where errors can occur. Some big errors may appear when equipment is broken or misused. These errors are rather rare. Despite such large errors we have to keep in mind all smaller errors which may appear and influence whole radiotherapy process and therefore would be considered as systematic errors. First of all, during CT, the tattoos are done on skin, which can move relative to internal anatomy (this is not the case for head and neck patients because they do not have tattoos[1]). Another source of error is related to the fact that moving organs and target are frozen during CT scan in some position, which does not have to be the most representative one. Third source of error lies in patient setup. All these errors mentioned above, which can appear during CT scan, have impact on whole radiotherapy process and therefore should be considered as systematic errors.

CT scans are used to calculate dose in target volume and normal tissues. Before it is done those structures have to be outlined on the CT scans. Delineation process also may lead to

[1]There are known situations that patient removed the adhesive tapes with the marker and afterwards put it in a different position causing a large positional error.

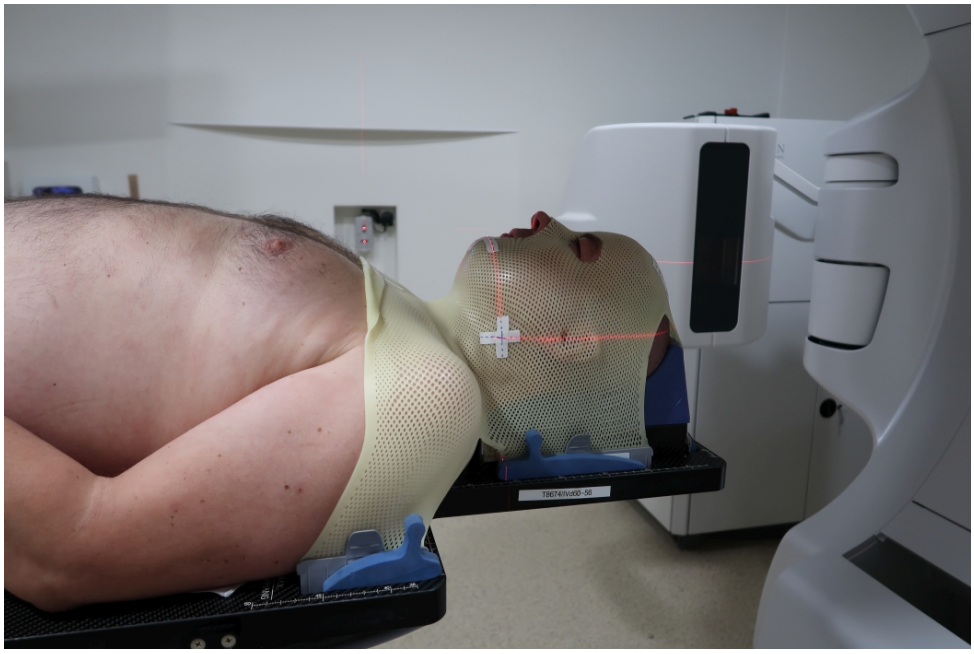


Figure 2.1: Photography presents a patient lying on a treatment couch in the individualised thermoplastic mask (yellow) for head and neck region immobilization. Laser system (red) crossing on adhesive tapes attached to the mask can be also seen. Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

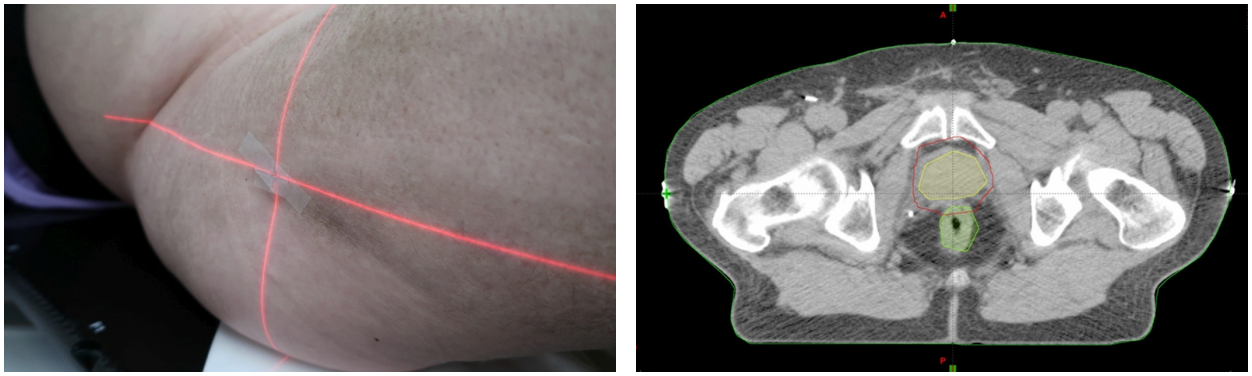
Figure 2.2: Photography presents a patient lying on a treatment couch in the individualised vacuum bag (blue) with additional knee support (green). Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

Figure 2.3: Laser system (red) in pelvic region. Marker can be seen as a small black dot on the laser crossing. Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

certain systematic errors. There are three sources of errors at that step of patient preparation to treatment. First effect is known as partial volume effect [52] and is related to restricted resolution in the axis perpendicular to the scans. In other words — last scan with the structure delineation does not have to be the end of structure in patient body. Next two sources of errors are related to intra- and inter-observer variations. The same observer asked to delineate the target volume twice would do it slightly differently [25]. Even though there are protocols defining how the target and structures should be delineated the delineation uncertainty would be even bigger if different observers are asked to delineate the same structure [25]. Due to the fact that the whole radiotherapy process depends on treatment plan which is based on CT and delineations made on it, delineation errors are systematic ones.

The last step of radiotherapy process is irradiation. In a typical treatment it is done during a course of treatment when patient receives many (usually 25-35) fractions of the whole treatment, typically one each day. Setup errors as well as organ and target motion can occur not only during treatment preparation but also during irradiation. It has to be clarified that some errors which lead to systematic errors during CT scan, during irradiation would have a random nature, because in each fraction they may be different. For example the organ motion frozen on a CT can lead to systematic error, while during radiotherapy it would lead to random inter- or intrafraction error.

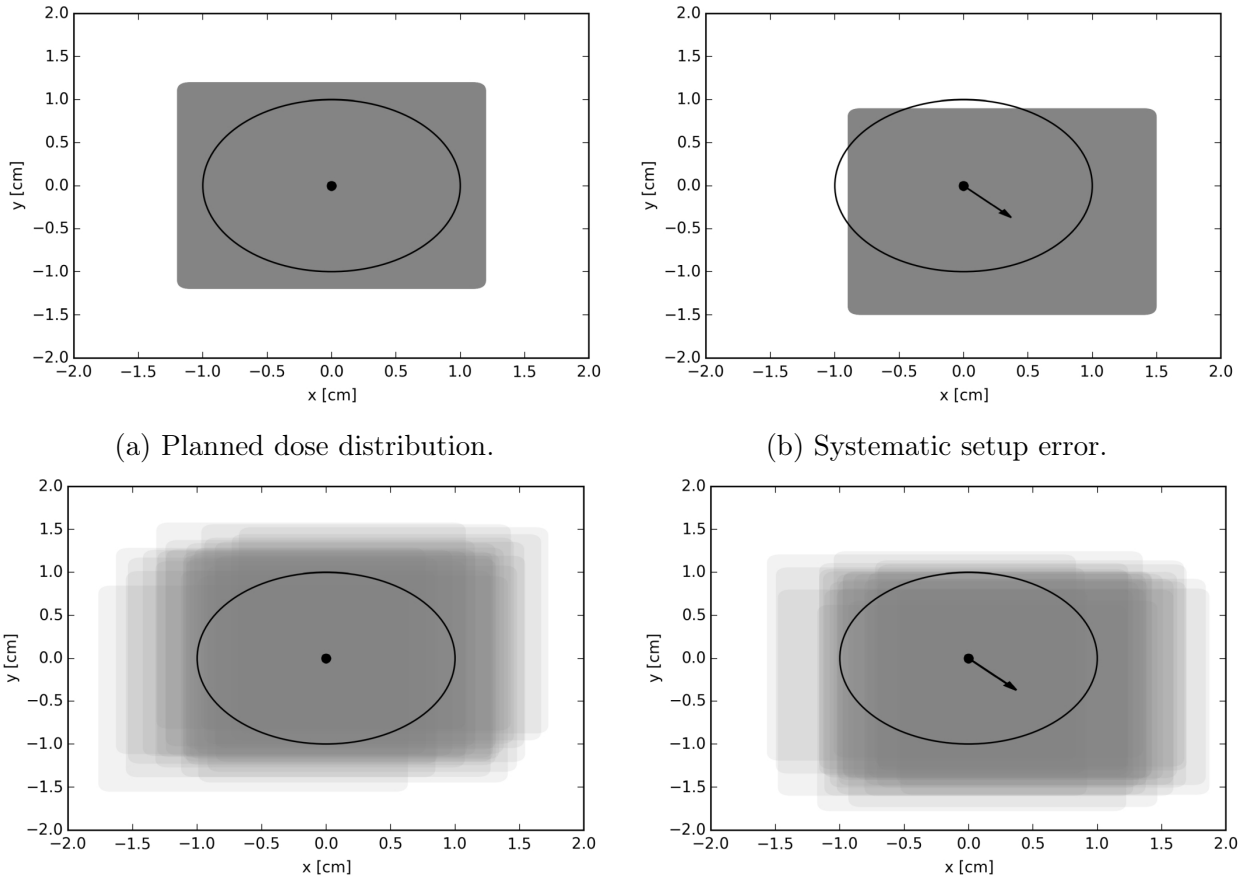
As was described before we divide the errors into the systematic (preparation) and random (treatment execution) ones. That division is important due to the fact that both kind of errors have a different impact on cumulative dose distribution [89]. Assume that we have planned the ideal i.e. homogeneous dose distribution around the target volume (see Fig. 2.5a). Systematic error would shift the dose distribution (see Fig. 2.5b) while the random errors would blur it (see Fig. 2.5c). In other words systematic error can cause underdosage in the part of target volume and random errors usually make dose distribution less homogeneous. Fig. 2.5d shows



(a) Markers placed on body.

(b) Markers seen on CT slice.

Figure 2.4: Methodology of transformation of laser crossing point into CT. Radio-opaque markers are placed on patient body prior to CT scan (left figure) in order to represent tattoos. These markers can be seen on CT exam slice (right figure). Photography was taken with patient permission in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.



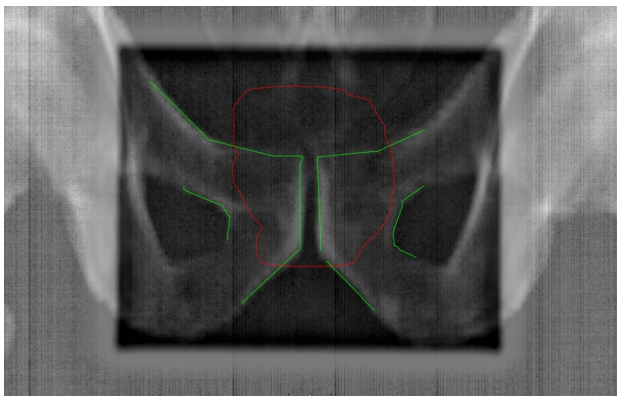
(a) Planned dose distribution.

(b) Systematic setup error.

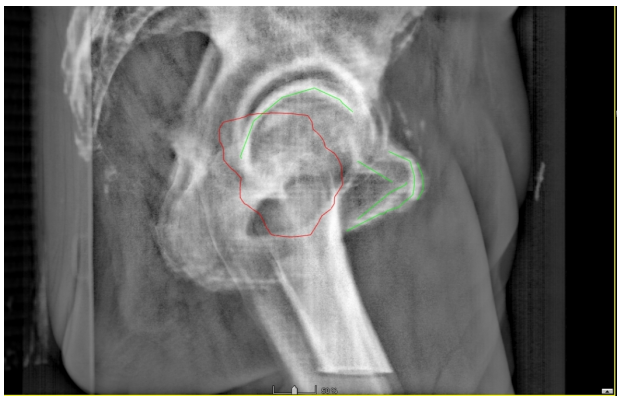
(c) Random setup errors.

(d) Systematic and random setup errors.

Figure 2.5: Influence of setup errors on dose distribution. The black circle indicates the target volume. The dot indicates the planned isocenter position. Vector shown on 2.5b and 2.5d indicates the systematic displacement of the treatment isocenter relative to the planned one.



(a) Comparison of anterior setup image (0°).



(b) Comparison of lateral setup image (270°).

Figure 2.6: Setup images performed prior to treatment aligned to DRRs on which bony structures are delineated. Images were acquired in Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

the typical case during the radiotherapy treatment i.e. the existence of both systematic and random errors.

There are different methods used to reduce the systematic and random errors. Former ones can be minimized by verification protocols which would be described further in section 2.3. Random errors can be minimized by improvement of patient immobilization.

The common way to detect, at least some errors, is to compare the Digitally Reconstructed Radiographs (DRRs), based on CT scans, with setup images taken prior to treatment when patient is already lying on the treatment couch (see Section 2.6). The methodology of such a comparison is based on delineation of important and meaningful structures on DRRs. Knowing the distance between the setup field isocenter and the structures one can define the exact position of isocenter on the setup image relative to the patient anatomy.

According to definitions proposed by Bijhold [4] the displacement vector $\boldsymbol{m_{p,f}}$ measured for patient p in fraction f can be given as:

$$\boldsymbol{m_{p,f}} = \boldsymbol{\Delta_p} + \boldsymbol{\delta_{p,f}} + \boldsymbol{E_p} + \boldsymbol{\varepsilon_{p,f}} \tag{2.1}$$

where:

$\boldsymbol{\Delta_p} \rightarrow$ systematic displacement between planned and treatment isocenter in relative to patient anatomy for patient p during whole treatment

$\boldsymbol{\delta_{p,f}} \rightarrow$ random deviation of displacement between planned and treatment isocenter position in relative to patient anatomy for patient p in fraction f

$\boldsymbol{E_p} \rightarrow$ systematic measurement error for patient p (the source of this error can be for example the error during the structure delineation on the DRRs or the error in definition of setup field edges which is normally used in order to define the treatment isocenter position on the image)

$\boldsymbol{\varepsilon_{p,f}} \rightarrow$ random measurement error which can vary from fraction to fraction (that error would account for inaccuracy of the alignment between the DRRs and setup images)

Equation (2.1) can be rewritten in order to emphasize its systematic and random component:

$$\boldsymbol{m_{pf}} = \boldsymbol{S_p} + \boldsymbol{\mu_{pf}} \tag{2.2}$$

where:

$\boldsymbol{S_p} = \boldsymbol{\Delta_p} + \boldsymbol{E_p}$

$\boldsymbol{\mu_{pf}} = \boldsymbol{\delta_{pf}} + \boldsymbol{\varepsilon_{pf}}$

Assuming that $\boldsymbol{S_p}$ and $\boldsymbol{\mu_{pf}}$ have normal distribution and are not correlated Bijhold [4] states that their distributions can be interpreted as covariance matrices.

Somehow alternative notation is used by van Herk [90]. Variation of systematic errors calculated for each patient p in population P is denoted as Σ. In order to make description more transparent I would focus only on one principal axis. Equations for all other axes are the same. Assume that we have measured $m_{p,f}$ for each patient p from the group of patients P who were undergoing the same kind of treatment (same cancer site, same positioning, similar fractionation and so on). If we have done setup measurements in all fractions f for each patient p, one can calculate the mean setup error $\overline{m_p}$ of patient p and its standard deviation SD_p:

$$\overline{m_p} = \sum_{f=1}^{F} \frac{m_{p,f}}{F} \tag{2.3}$$

$$SD_p = \sqrt{\sum_{f=1}^{F} \frac{(m_{pf} - \overline{m_p})^2}{F-1}} \tag{2.4}$$

where:

$F \rightarrow$ total number of fractions in a treatment course

The mean setup error $\overline{m_p}$ describes the systematic error for patient p, while the standard deviation SD_p describes the magnitude of random errors for that patient.

The systematic error for the whole population P is described with the mean of all means M (i.e. the mean systematic error for whole population) and its standard deviation Σ:

$$M = \sum_{p=1}^{N} \frac{\overline{m_p}}{N} \tag{2.5}$$

$$\Sigma = \sqrt{\sum_{p=1}^{N} \frac{(\overline{m_p} - M)^2}{N-1}} \tag{2.6}$$

where:

$N \rightarrow$ number of patients p in their population P

According to van Herk [89] Σ describes how well, i.e. how reproducible, the treatment preparation is performed.

In notation used by van Herk a term σ describes random errors by providing the best possible estimate of mean SD_p in population of patients. As van Herk [89] mentioned, for limited number of fractions with measurement it is hard to prove that differences between patients exist. Therefore in order to describe the group mean of SD_p the Root Mean Square (RMS) should be used:

$$\sigma = \sqrt{\sum_{p=1}^{N} \frac{{SD_p}^2}{N}} \tag{2.7}$$

The width of SD_p distribution can be calculated as follows [91]:

$$SD_{SD} = \sqrt{\sum_{p=1}^{N} \frac{(\overline{SD_p} - SD_p)^2}{N-1}} \tag{2.8}$$

In practice it is rather impossible to minimize setup errors to zero value. Therefore in order to take them into account during radiotherapy planning the margin is added to the target volume - for details on how it is done please look into section 2.2.

2.2 CTV-PTV margins

In order to standardize the nomenclature and procedures used in radiotherapy the International Commission on Radiation Units and Measurements (ICRU) provided the Report No. 50 "Prescribing, recording and reporting photon beam therapy" [47] and then its supplement — Report No. 62 [92]. The latter one gives a three-volume concept which describes irradiated target volume in a detailed manner [9, 78] — see Fig. 2.7.

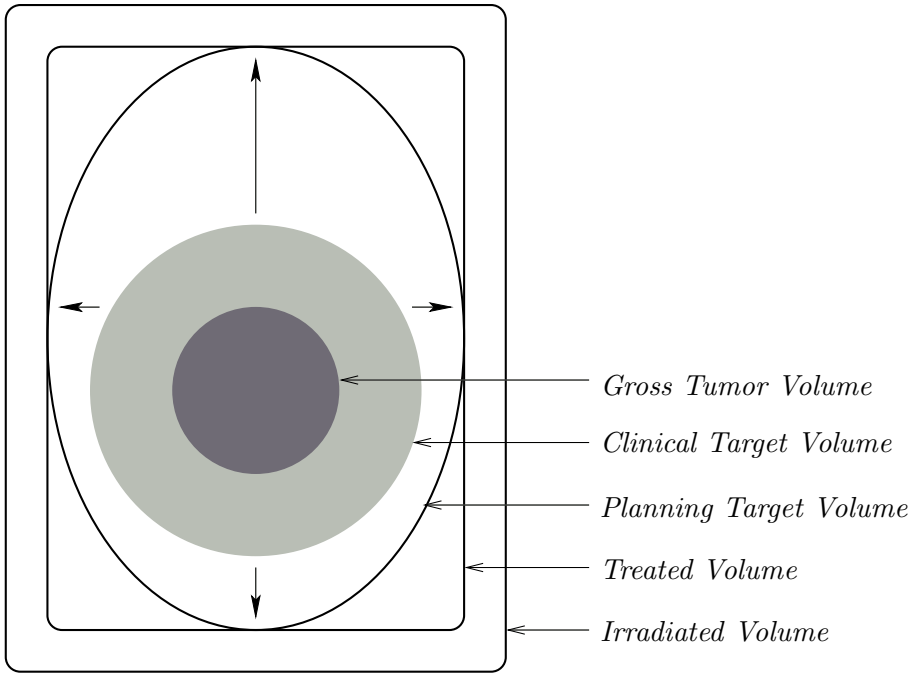


Figure 2.7: Schematic illustration of volumes used in prescribing and reporting radiotherapy (graph derived from my Master Thesis [32]).

Gross Tumor Volume (GTV) also known as Gross Target Volume consist of tumor volume i.e. the volume with macroscopic extent of the tumor. It can be detected e.g. on diagnostic scans like CT, Positron Emission Tomography (PET) or Nuclear Magnetic Resonance (NMR) (known also as Magnetic Resonance Imaging (MRI)).

The definition of Clinical Target Volume (CTV) given by ICRU [47] is: "The CTV is a tissue volume that contains a gross tumor volume (GTV) which is the gross palpable or visible/demonstrable extent and location of the malignant growth, and/or subclinical microscopic malignant

disease, which has to be eliminated. This volume has to be treated adequately in order to reach the aim of therapy: cure or palliation." GTV can be defined based on imaging (like CT or MRI) or palpation examination. GTV border is clear and well defined. It has to be emphasized that GTV cannot be defined after R0 (i.e. no cancerous cells in margin) surgery. In the presence of GTV, CTV is created usually by adding isotropic margin. If GTV, as a result of surgery or difficulty in tumor recognition on CT/MRI (like in prostate case), does not exist the CTV is delineated itself.

The definition of Planning Target Volume (PTV) given by ICRU [47] is: "The PTV is a geometrical concept, and it is defined to select appropriate beam sizes and beam arrangements, taking into consideration the net effect of all the possible geometrical variations and inaccuracies in order to ensure that the prescribed dose is actually absorbed in the CTV." The PTV is created by adding margin to CTV. That margin does not have to be isotropic because errors in different directions can be of different magnitude.

ICRU Reports define also the Treated Volume (TV) i.e. the volume receiving the prescribed dose (usually 95% of prescribed dose) and Irradiated Volume (IV) i.e. the volume included in an isodose with a possible impact on normal tissues.

There are many ways of calculating the margin between CTV and PTV [89]. The two most popular formulas were given by van Herk [90] and Stroom [82]. The first one, as the most often used in clinical practice, would be described in a more detailed way below.

2.2.1 van Herk formula for CTV-PTV margin calculation

Marcel van Herk et al. proposed their formula for CTV-PTV margin calculation in the year 2000 [90]. This concept is based on probability histograms of the cumulative dose over a population of patients (dose-population histograms). These histograms are derived with analysis of the blurred dose distribution (see Fig. 2.5c), i.e. the dose distribution including random errors. Probability of specified target coverage in population of patients is checked and presented as dose-population histograms. Assume, for point CTV, that dose blurring is the same for whole patient population. That blurring of cumulative dose distribution D_{cum} is done by incorporation of all random errors i.e. organ motion and setup error. Let further assume that we know the probability density function of systematic errors. Then we can calculate (by integration) how many systematic errors and how many patients (each patient have one systematic error) would have the cumulated dose (the blurred one) above the specified threshold ($D_{threshold}$). Afterwards we can prepare the histogram of probability in patient population in case of the CTV receiving at least a given dose. In other words we can look into the blurred dose distribution to check the volume surrounded by the isodose of $D_{threshold}$. Finally we check on the probability density distribution ($P(D_{cum} > D_{threshold})$) of how many patients would fall into that volume. It is the same as checking percentage of possible systematic errors which are within that volume.

This can be written in equations as follows (bold letters describe vectors):

$$\mathrm{C} : \{\mathbf{x} | D_{cum} > D_{threshold}\} \tag{2.9}$$

$$P(D_{cum} > D_{threshold}) = \int_C Q(\mathbf{z}) d\mathbf{z} \tag{2.10}$$

where:

$D_{cum} \rightarrow$ cumulative dose in CTV, i.e. dose which was blurred in order to consider random errors

$D_{threshold} \rightarrow$ given value of dose which describes requirement for example for minimum dose in CTV

$\mathbf{C} \rightarrow$ collection of all possible positions x of CTV in which the condition of $D_{cum} >= D_{threshold}$ is fulfilled

$\mathbf{z} \rightarrow$ systematic error

$Q(\mathbf{z}) \rightarrow$ probability density of $\mathbf{z}$

Assumption of no correlation between different errors (organ motion, delineation error and setup error) leads to conclusion that $Q(\mathbf{z})$ can be expressed as a normal (Gaussian) distribution with zero mean and variance $\mathbf{\Sigma}$ which would be sum of variances of each kind of error mentioned above.

$$Q(\mathbf{z}) = G(\mathbf{z}, \mathbf{\Sigma}^2) = G(\mathbf{z}, \mathbf{\Sigma_m}^2 + \mathbf{\Sigma_s}^2 + \mathbf{\Sigma_d}^2) \tag{2.11}$$

where:

$\mathbf{\Sigma_m}^2 \rightarrow$ variance of the organ motion during treatment preparation

$\mathbf{\Sigma_s}^2 \rightarrow$ variance of the setup error during treatment preparation

$\mathbf{\Sigma_d}^2 \rightarrow$ variance of the delineation error

The creation of margin formula is a reversion of the procedure of preparing the dose-population histograms. At first one has to decide on the desired probability that patient from the population would receive at least the specified dose in CTV. Let us take as an example a probability of patient population equal 90% and minimum required dose in CTV equal 95% of prescribed dose. At first we have to choose all possible systematic errors which would fall into 90% of their probability density distribution. The extent of these systematic errors can be easily written as $\alpha\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is a vector, as it can be different in all directions, of all systematic errors components (organ motion, setup error and delineation error). For 3D case and probability/prescribed dose levels mentioned above value of α would be 2.5. As it is highlighted by van Herk this is purely geometrical concept and therefore it does not depend on dose distribution and the shape of CTV.

Second step is to find an extra margin in order to incorporate the random errors. It means that necessary extra extend of the CTV has to be found in order to reassure that 95% isodose of blurred dose distribution would surround all the points of CTV even though they would change location due to random errors. If we would describe the distance between the 95% and 50% isodose of blurred dose distribution, $D_{blurred}$, as $\beta\boldsymbol{\sigma}$ and distance between the 95% and 50% isodose of planned dose distribution, $D_{planned}$, as $\beta\boldsymbol{\sigma_p}$ then the extent in CTV can be written as $\beta\boldsymbol{\sigma} - \beta\boldsymbol{\sigma_p}$.

In the end we come to the total PTV margin recipe:

$$\mathbf{M^{PTV}} = \alpha\mathbf{\Sigma} + \beta\boldsymbol{\sigma} - \beta\boldsymbol{\sigma_p} \tag{2.12}$$

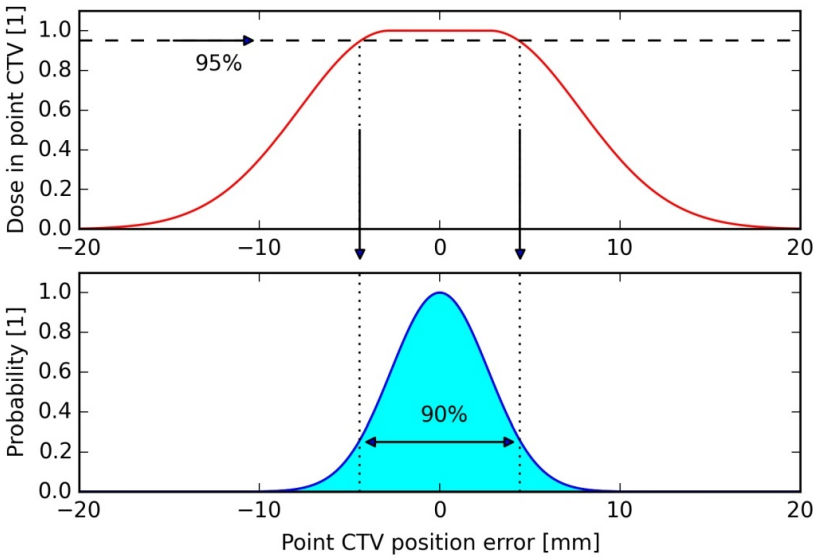
where:

$\boldsymbol{\sigma}^2 = \boldsymbol{\sigma_m}^2 + \boldsymbol{\sigma_s}^2 + \boldsymbol{\sigma_p}^2$ and:

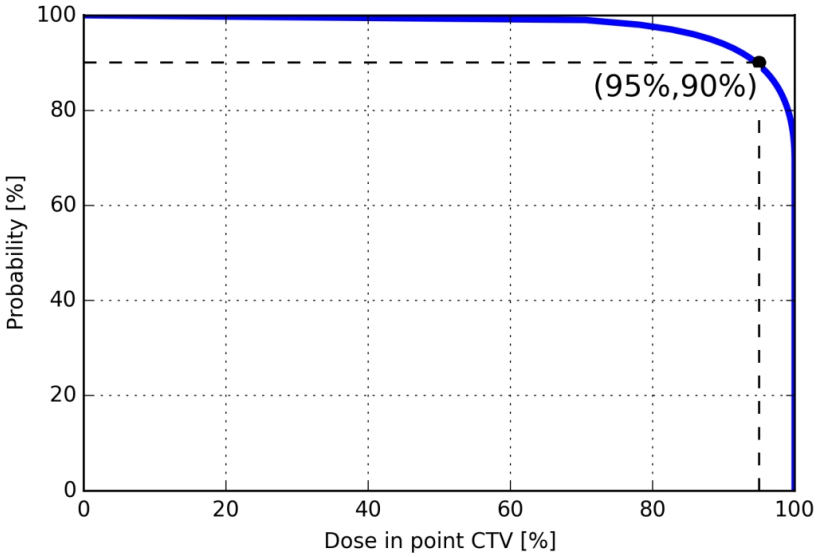
$\boldsymbol{\sigma_m}^2 \rightarrow$ variance of the organ motion during treatment execution

$\boldsymbol{\sigma_s}^2 \rightarrow$ variance of the setup error during treatment execution

$\boldsymbol{\sigma_p} \rightarrow$ beam penumbra width



(a) Derivation of dose-population histogram.



(b) Dose-population histogram.

Figure 2.8: Figure presents the methodology of creating probability histogram of the cumulative dose over a population of patients (dose-population histogram). For simplicity CTV consisting of one point is assumed. Fig. 2.8a shows how dose-population histogram is obtained. Red curve on upper graph presents dose in point CTV in relation to error in CTV position. We have to decide on $D_{threshold}$ which should be delivered (e.g. the minimum dose in CTV, here 95%). Then we look into the range/collection of all possible CTV positions in which the dose would be at least $D_{threshold}$. Knowing this collection we check what fraction of whole population it is (in other words we check the probability that CTV would have this location) — see lower graph of Fig. 2.8a for probability of CTV displacement (CTV position error). That procedure leads us to a single point in dose-population histogram (Fig. 2.8b) and therefore has to be repeated for each $D_{threshold} \epsilon [0\%; 100\%]$. Dose-population histogram prepared for data shown here is presented in Fig. 2.8b. Selected point (95%, 90%) shows that 95% of dose would be achieved in 90% of patients. These figure was created on the basis of van Herk paper [90].

In Eq. (2.12) margin $\mathbf{m_{PTV}}$ and all variances are vector variables. Therefore also non-isotropic margins can be calculated with the usage of this formula. Equation (2.12) can be rewritten if the penumbra width has to be excluded. In that case we end up with:

$$\mathbf{M^{PTV}} = \alpha\boldsymbol{\Sigma} + \kappa\boldsymbol{\sigma'} \tag{2.13}$$

where:

$\boldsymbol{\sigma'} = \sqrt{\boldsymbol{\sigma_m}^2 + \boldsymbol{\sigma_s}^2}$ and:

$\kappa \rightarrow$ parameter achieved for linear fit used to approximate $\beta\boldsymbol{\sigma} - \beta\boldsymbol{\sigma_p}$ with $\kappa\boldsymbol{\sigma'}$ for SD of random errors in range [0 mm, 5 mm].

Parameter κ is valid only for limited range of uncertainties and for a specified penumbra width. Setting the $\boldsymbol{\sigma_p}$ to 3.2 mm (which is adequate for 5 mm distance between 50% and 95% isodose in normal, i.e. not blurred dose distribution) and setting the requirement of 95% minimum dose in CTV in 90% of patients we come to the most applied margin formula [89] (please not that often $\boldsymbol{\sigma'}$ is denoted as $\boldsymbol{\sigma}$):

$$\mathbf{M^{PTV}} = 2.5\boldsymbol{\Sigma} + 0.7\boldsymbol{\sigma'} \tag{2.14}$$

2.3 Verification Protocols

The aim of radiotherapy is to deliver high, generally uniform, dose to target volume and spare healthy tissue at the same time. The magnitude of irradiated healthy tissue volume depends on the size of CTV-PTV margin. Therefore it is important to minimize that margin. As was previously described (see Section 2.1) during radiotherapy one has to deal with two types of errors: random and systematic ones. Reduction of random errors can be achieved with better immobilization and on-line verification procedure [94]. In such a procedure setup images are taken prior to each fraction. There are compared with reference images immediately (on-line) and corrections are applied. Patient is irradiated afterwards. That makes this procedure rather time-consuming. The on-line verification protocol seems to be very efficient as it would correct for daily setup error. However it was shown that intrafraction motion might have greater impact in that case [57].

Reduction of systematic errors can be done by off-line verification protocols (OVP). The main idea of OVP is to estimate patient systematic error and correct for it. This estimation would not be perfect and therefore some residual systematic error would remain. Still, if OVP works accurately variation of residual systematic error for whole patient population — Σ_{res} would be smaller than Σ. Due to van Herk formula (Eq. (2.14)) decrease in systematic errors is more important for margin reduction than decrease in random errors.

2.3.1 No Action Level Protocol

No Action Level (NAL) protocol was proposed by de Boer and Heijmen in 2001 [14]. It is widely used in the clinical practice [17, 54, 76] and some extensions were proposed [16, 15, 55]

NAL protocol needs a minimum workload. In that protocol each patient has setup images done in first n fractions. During this first n fractions no corrections are applied and a comparison of setup images with reference images is done after the treatment (i.e. offline). After n fractions the average setup vector $\boldsymbol{V_n}$ is calculated. It consist of average setup error in each direction. All

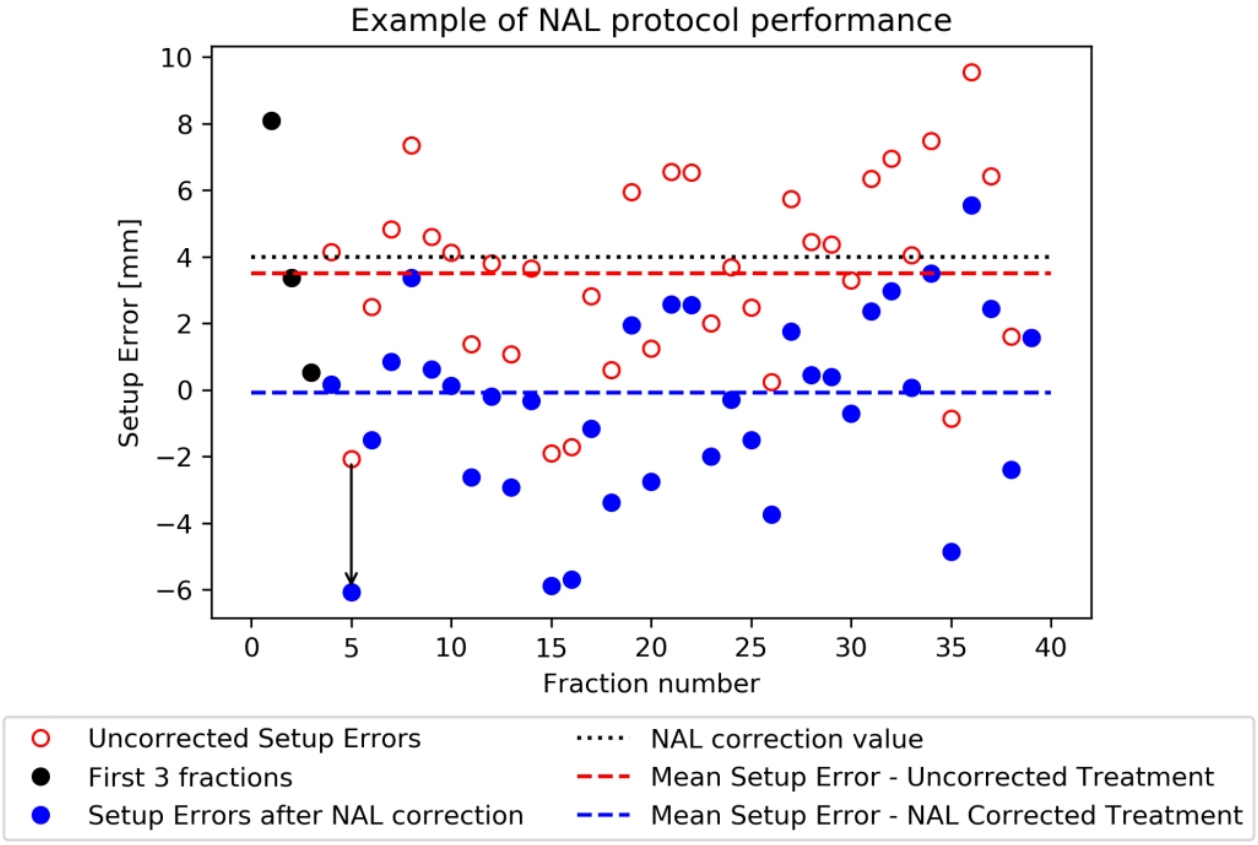


Figure 2.9: Example of NAL protocol performance. Figure shows NAL in case when mean setup error (i.e. the correction value) is well estimated on the basis of 3 first fractions. Arrow shows an example shift between corrected an uncorrected Setup Error.

upcoming fractions are corrected with $-\boldsymbol{V_n}$ — see Fig. 2.9. Value of n should be large enough to sufficiently estimate systematic error and small enough to assure minimal workload. De Boer and Heijmen [14] suggested $n = 3$, Bortfled [7] concluded that it would be more efficient to use $n = 4$.

2.3.2 Extended No Action Level Protocol

The much applied NAL protocol (see Section 2.3.1) was extended by de Boer and Heijmen in 2007 in order to deal with inter-fraction time trends [16]. In first three fractions this protocol works in the same manner as NAL protocol, so images are taken and no correction is applied. Afterwards, the mean setup error is calculated and taken as a correction value for upcoming week of treatment (five fractions). At the last fraction of each following week images are also taken. A linear fit is done for uncorrected setup errors and new correction value is calculated according to this linear fit. In such a way correction value is updated weekly. An example of the extended NAL (eNAL) protocol performance is shown in Fig. 2.10.

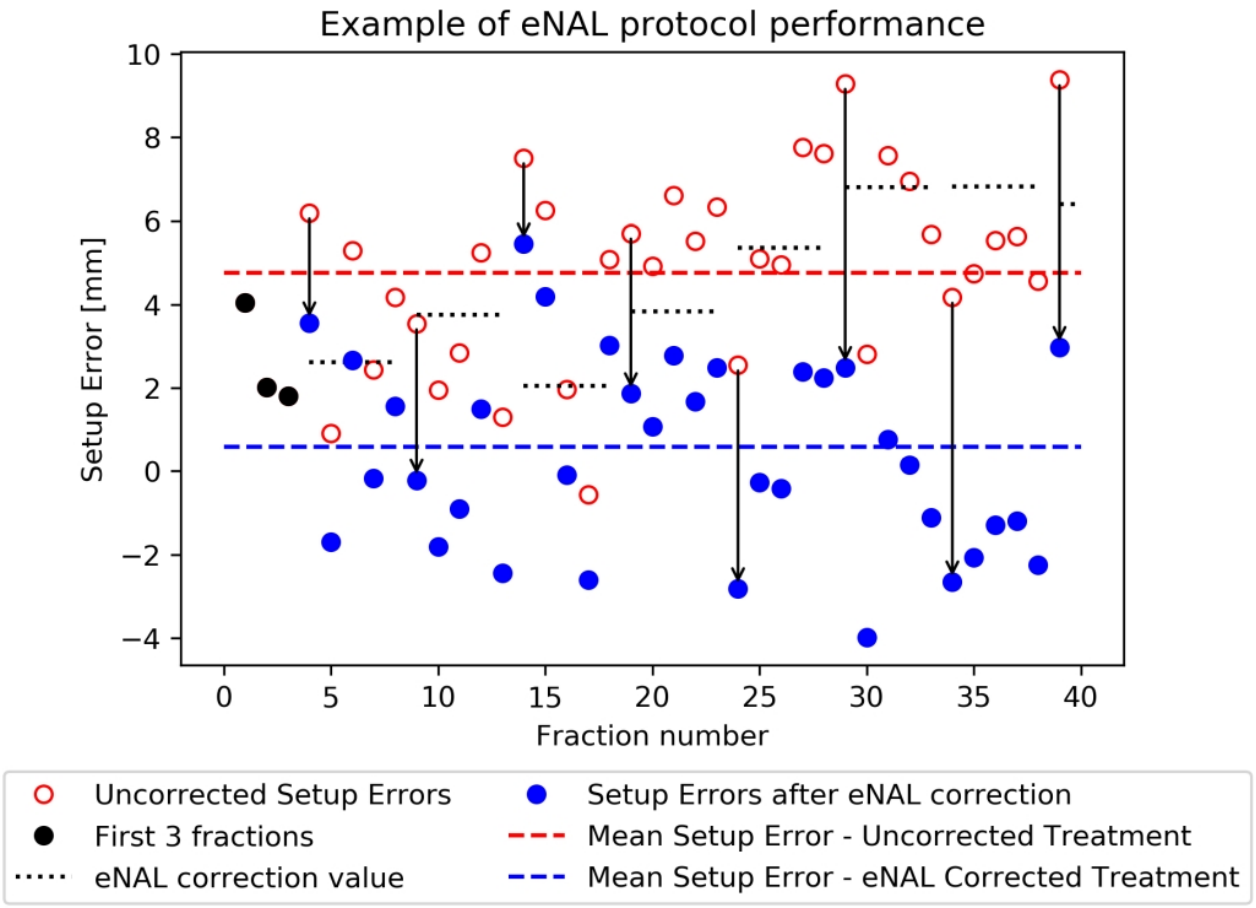


Figure 2.10: Example of eNAL protocol performance in presence of inter-fraction time trend. First three fractions (black dots) are irradiated without any correction. Afterwards a correction is calculated and applied for five upcoming fractions (first arrow indicates this correction value). In 8th fraction an imaging is performed and a linear fit is done for uncorrected setup errors. On the basis of this fit correction the value for upcoming five fractions is recalculated. The imaging is done in each fifth fraction, followed by a linear fit and update of correction factor. Arrows show first fractions after the new correction value was calculated and dotted black lines show the correction value and fractions in which it was applied. It can be seen how the correction value is changing in time.

2.4 Dose Volume Histograms (DVHs)

After the radiotherapy plan is created in TPS, it has to be accepted by a physician. Only after that acceptance the further preparation (double check of dose calculations, pretreatment QA, transfer to patient management and/or record and verify system) and finally irradiation may happen. In order to check the plan the physician examines dose distributions in single CT scans. The conformity, hot and cold spots are investigated visually. As there are some dose-volume constraints in the planning protocols they also have to be checked. It is done with dose volume histograms (DVH).

There are two types of DVHs: differential and cumulative. Both are representations of dose distribution but without the spatial information. Creation of both types of histograms is explained below. For simplicity the 2D dose distribution presented in Fig. 2.11 was used. Knowing dose matrix (i.e. dose distribution) one has to check which voxels lie in the structure. Coping with voxels partly in the structure is done differently in different TPS. In the example I consider all voxels with PTV. In classical histogram one has to calculate number of occurrences (see Fig. 2.12a). As it would be hard to present in TPS (there are many fluctuations and bins are usually very thin) the decision was made to take only middle position of each bar and connect them with the line (see Fig. 2.12b). Ideal uniform distribution in PTV should have Dirac delta shape in differential DVH. Let us denote a point on a differential histogram as DVH_{diff}. In order to present data in a more clear and understandable way cumulative histogram DVH_{cum} is created:

$$\mathrm{DVH}_{cum}(D_{i+1}) = \left(1 - \int_0^{D_i} \mathrm{DVH}_{diff}(D_i)\right) \cdot V_{total} \tag{2.15}$$

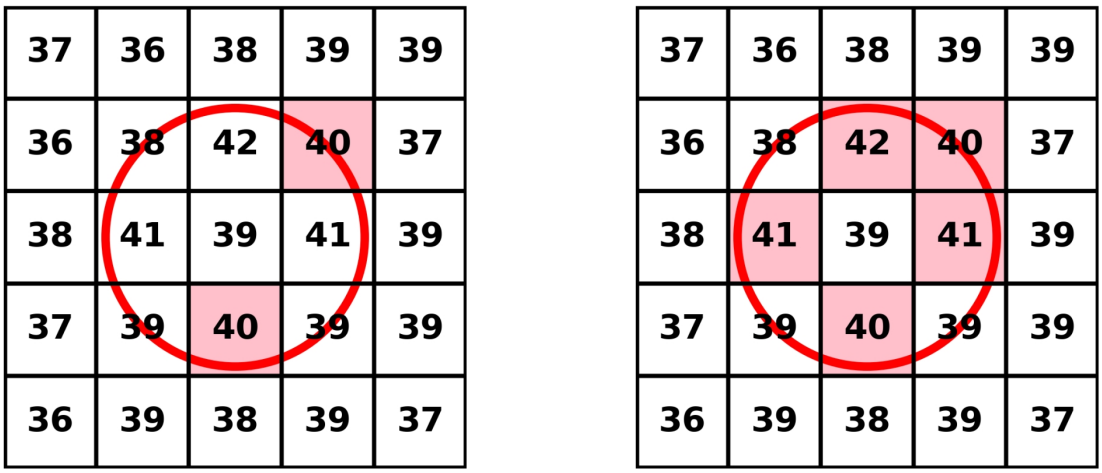
where V_{total} is total structure volume. The cumulative histogram created for presented example is shown in Fig. 2.13.

2.4.1 Dose Constraints

Statistical dose evaluation is done on basis of some dose or dose-volume parameters (ofted called dosimetric parameters). Usually dose-volume parameters are presented as D_V or V_D. First parameter D_V (e.g. $D_{2\%}$ or D_{1cc}) is such a dose D that doses greater or equal D would be in volume equal V (see Fig. 2.14a). Second parameter V_D (e.g. $V_{20} = V_{20\,\mathrm{Gy}}$) is a volume V in which doses would be equal D or greater (see Fig. 2.14b).

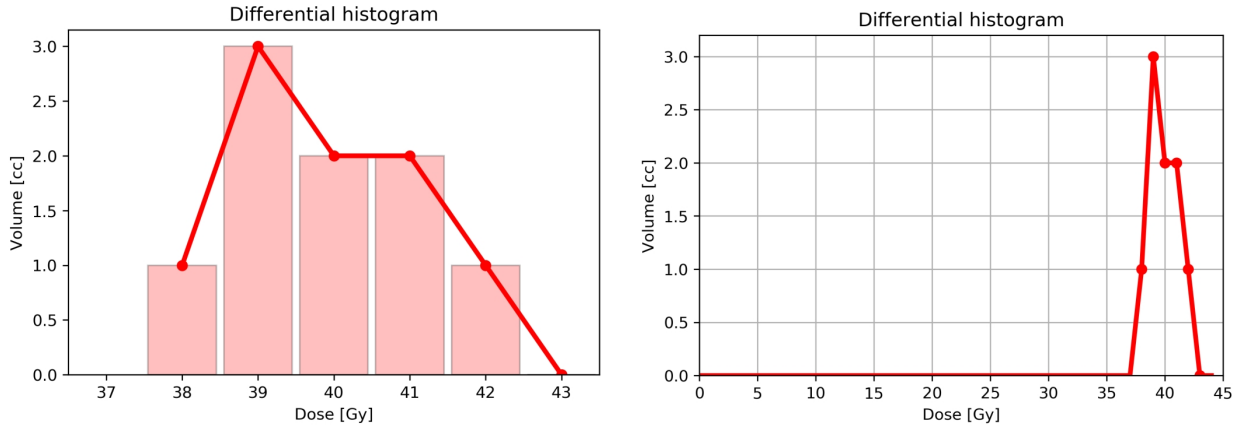
For target volumes maximum dose D_{max}, minimum dose D_{min} and mean dose D_{mean} are usually evaluated [47]. Due to the fact that D_{max} and D_{min} are point doses and they depend on structure delineation and dose grid the concept of dose close to minimum $D_{98\%}$ and dose close to maximum $D_{2\%}$ were proposed [18].

Dose parameters evaluated for Organs at Risk (OARs) depend on the structure of the organ. For so-called serial OAR for which destruction of one element can cause destruction of whole organ maximum dose D_{max} is evaluated. For so-called parallel OAR for which even if some part is destructed other part will remain working D_{mean} and V_D or D_V parameters are evaluated. For OARs with mixed structure all of these parameters are taken into consideration.



(a) Differential (b) Cumulative

Figure 2.11: Simple 2D dose matrix with a PTV contour (red circle) is presented. While considering $D_i = 40$ Gy, for a differential histogram only two voxels (marked in light red) would be taken into account. For a cumulative histogram, while considering $D_i = 40$ Gy, all voxels with $D \geq 40$Gy (marked in light red) would be counted.



(a) Classical histogram (b) Differential DVH

Figure 2.12: Explanation on how differential DVH is created. The classical histogram is created from the dose matrix at first (2.12a). Afterwards the middle positions of bars are taken and joined with line (2.12b).

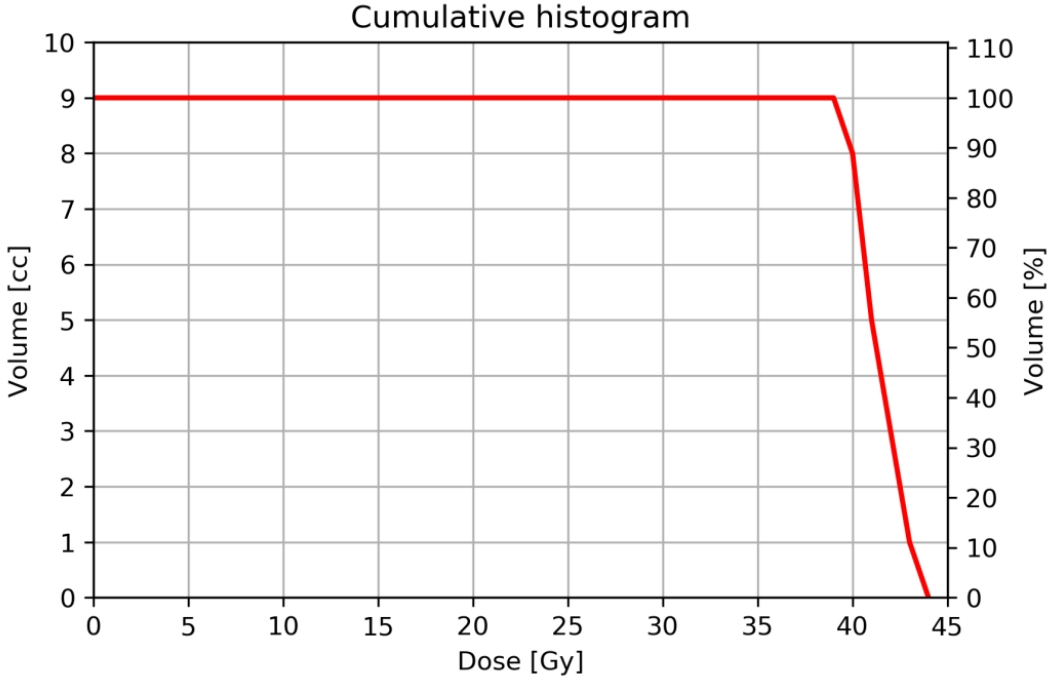
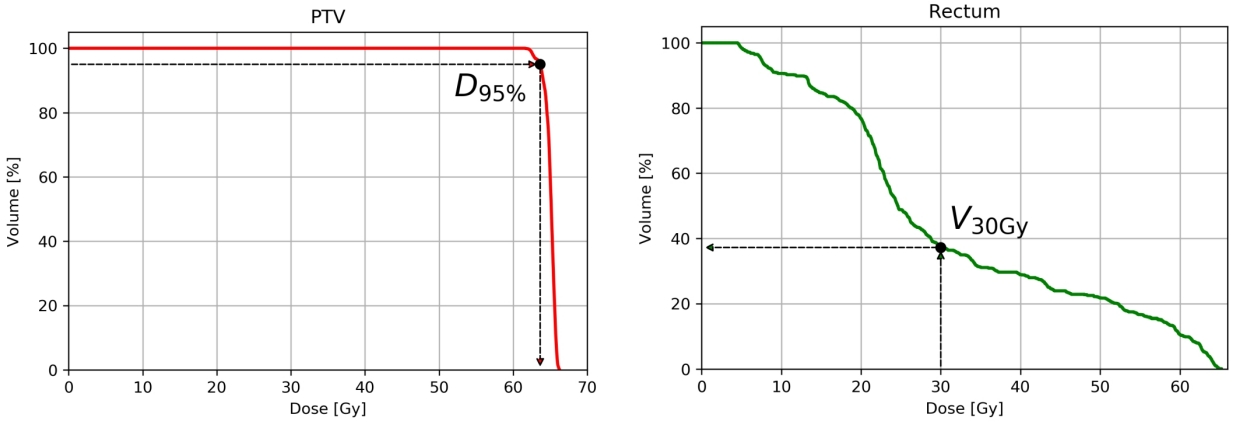


Figure 2.13: Cumulative Dose Volume Histogram. For some dose-volume constraints cubic centimeters are used, more often volume is given in percentage of total structure volume.



(a) D_V parameter

(b) V_D parameter

Figure 2.14: Illustration of dosimetric parameters. In order to check $D_V = D_{95\%}$ we look on volume $V = 95\%$ on the y-axis, search for point on a curve and check adequate dose D (Fig. 2.14a). In order to check $V_D = V_{30\text{ Gy}}$ we look on dose $D = 30$ Gy on the x-axis, search for point on a curve and check adequate volume V (Fig. 2.14b).

Chapter 3

Clinical Data

3.1 Background

Work presented in the thesis was done on the basis of two databases of setup errors for prostate cancer patients. First group of patients were irradiated in Maria Skłodowska-Curie Cancer Center and Institute of Oncology in Warsaw (COI). Second group of patients were irradiated in Erasmus MC University Medical Center Rotterdam (Erasmus MC).

Prostate cancer is one of the most frequent cancer case in males population for many industrialized nations [80, 84]. In Poland, in terms of frequency, prostate cancer has a second place (1st place goes to lung cancer) [20]. Prostate cancer can be treated with surgery, radiotherapy, hormonotherapy or with combination of these therapies [39, 40].

In current chapter I will describe both clinical databases in terms of underwent treatment and recorded setup errors. I will also provide some statistical analysis of the data.

3.2 Materials

3.2.1 Database from Maria Sklodowska-Curie Institute — Oncology Center in Warsaw (COI)

In COI patients suffering from prostate cancer can receive radiotherapy as a part of their treatment. There are two main schemes of the irradiation. First is a standard therapy: 25 fractions with 2.6 Gy per fraction. Second one is Simultaneous Integrated Boost (SIB) therapy [2, 21, 48, 49] consisting of 27 fractions with 2.6 Gy per fraction in CTV and 2.45 Gy per fraction in PTV. Standard therapy is usually done with 3D Conformal Radiotherapy (3D-CRT) or Intensity Modulated Radiation Therapy (IMRT) [65, 98] technique, while SIB treatment is mainly performed with IMRT or Volumetric Arc Therapy (VMAT) [8, 70, 97].

Prostate cancer patients treated in COI undergo special verification protocol which was published in 2012 [74]. Patient has to empty his bladder and drink 0.5 l of water. CT scan or irradiation is done half an hour later with full bladder. That leads to the same stability of prostate position as for empty bladder irradiation with the reduction of bladder volume receiving high dose [73]. Patient is treated in supine position with a knee support used for more relaxed position — see Fig. 3.1. During CT tattoos are done on skin to mark lasers intersection and radio-opaque materials are put at that places. The width of CT scans is 2.5 mm.

Figure 3.1: CT scan for prostate cancer patient. Knee support is shown. Photography was taken with patient permission in the Maria Sklodowska-Curie Institute — Oncology Center in Warsaw.

In the treatment room patient is set up in the treatment couch using tatoos and laser system. During first 3 fractions orthogonal setup images (0°and 90°— see Fig. 2.6) and Cone-Beam CT (CBCT) images are taken (see Fig. 1.4b). Afterwards the displacement vector between bony structures (defined on orthogonal images) and prostate gland (defined on CBCT) is calculated. That vector is assumed to be a constant displacement in all fractions and used as a correction factor. Therefore in all upcoming fractions orthogonal images are taken, patient is shifted on the basis on bony anatomy match and afterwards correction based on displacement between bony anatomy and protate gland is applied.

The retrospective look on the setup images performed and recorded before each fraction allowed to calculate setup error in the regard to isocenter (i.e. in the regard to the laser coordinate system). That was a raw setup error without implemented corrections. In this work the analysis of such setup error was done for 100 patients who received standard 3D-CRT treatment between the May 2013 and the August 2015.

Before each of 25 fractions, and also prior to corrections, orthogonal setup images were taken. They were compared to DRRs on the basis of bony anatomy [43].

Comparison between DRRs and setup images resulted in setup errors values in three dimensions. In the vertical setup image the Head-Feet (z axis) and Left-Right (x axis) setup error were defined. In the lateral setup image the Head-Feet and Anterior-Posterior (y axis) setup error were defined. The accuracy of determination of setup error was 1 mm. Head-Feet setup error was defined from both images and therefore its accuracy was 0.5 mm. Rotations were not included in the analysis.

3.2.2 Database from Erasmus MC University Medical Center Rotterdam (Erasmus MC)

The database of 835 prostate cancer patients who underwent 39 fractions radiotherapy in Erasmus MC was also evaluated. Setup images were taken in each fraction and compared with reference images (DRRs or CT) on the basis of implanted gold fiducials [61].

3.3 Methods

3.3.1 Testing normality

The Shapiro-Wilk test [79], which is considered to be the most powerful test for normality [30, 38], was used in order to check the hypothesis that setup errors of each particular patient for each axis have normal distribution. The null hypothesis of these tests were that the setup errors have normal distribution. The hypothesis that the distribution of mean patient setup error $\overline{m_p}$ and patient's standard deviation of setup errors SD_p have normal distribution in the entire population of patients was also checked.

3.3.1.1 Reliability of Shapiro-Wilk test

At first I checked whether 25 samples, which correspond to 25 fractions of treatment, are enough to acquire a reliable result of Shapiro-Wilk test. The check was done on the basis of simulation. 25 samples were randomized from a normal distribution with standard deviation calculated as RMS of clinical patient data for each axis. The exact values were: 2.0 mm for left-right, 3.0 mm for anterior-posterior and 1.6 mm for head-feet. Then I checked the p-value of Shapiro-Wilk test. Procedure was done 10^5 times for each axis. Histogram plots were prepared.

The results obtained in simulation are presented in Fig. 3.2-3.4. It can be seen that distribution of p-values is rather uniform and that number of counts with $p < 0.05$ corresponds to the significance level which was set to 5%.

3.3.1.2 Data analysis for single patients

The Shapiro-Wilk test was performed for each patient and each axis of setup error. I collected all p-values and looked into their distribution in order to check the ratio of patients with not normal setup error distribution. Histogram plots of setup errors for each patient and each error direction (left-right, anterior-posterior, head-feet) were prepared. Additionally I prepared histograms of p-values for each direction for all patients.

Due to the fact that not normal distribution in at least one axis can influence correction procedure or margin calculation procedure I was also interested in number of patients with normal distribution in all directions. Adequate analysis was performed.

3.3.1.3 Data analysis for population of patients

Ditribution of patient's mean setup error — $\overline{m_p}$ (see Eq. 2.3) was analyzed. Shapiro-Wilk test was used in order to check probability that data came from the normal distribution. Similar analysis was done for distribution of patient's standard deviation — SD_p (see Eq. 2.4). All principal directions were treated separately.

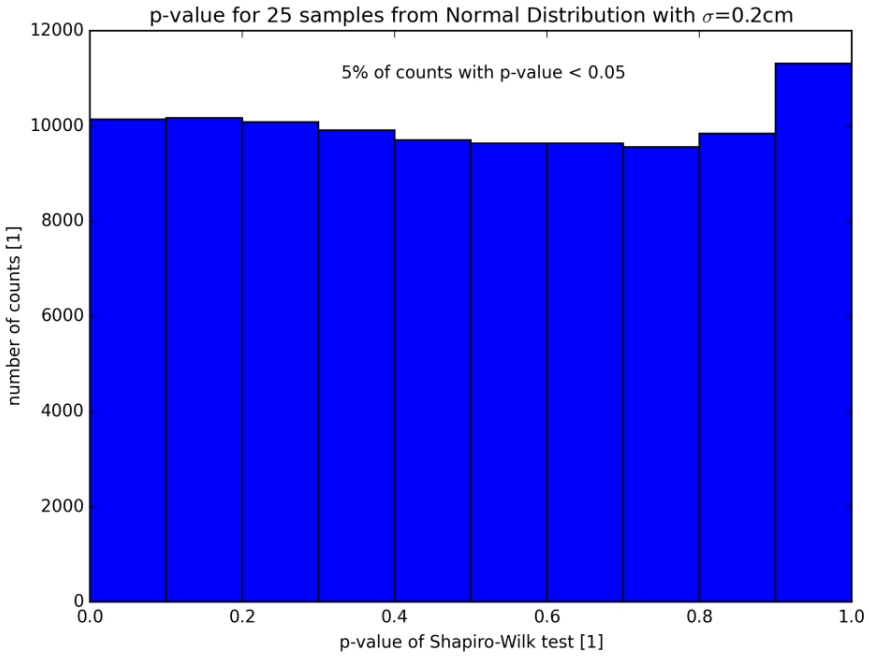


Figure 3.2: Histogram of p-values for setup error in left-right direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation.

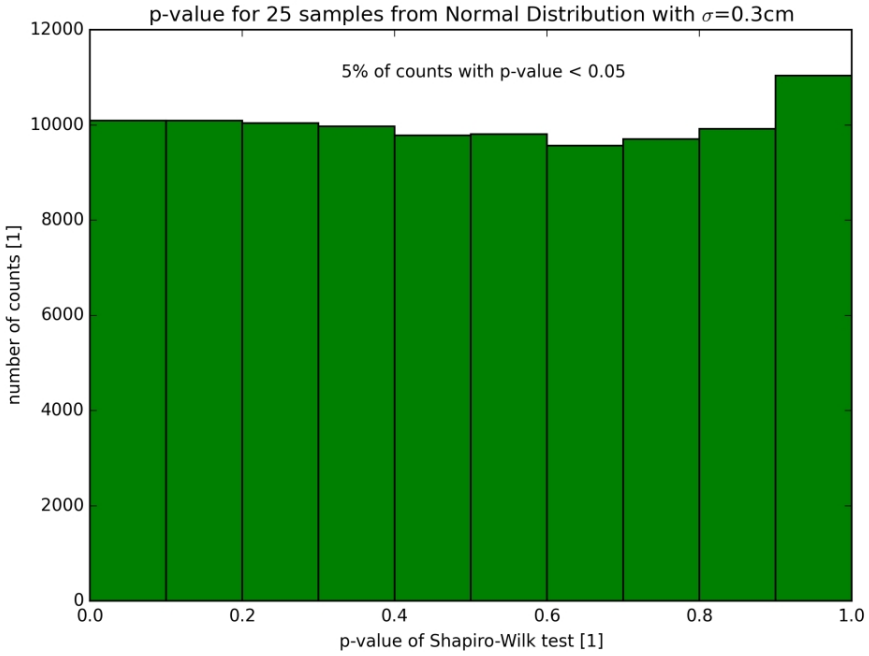


Figure 3.3: Histogram of p-values for setup error in anterior-posterior direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation.

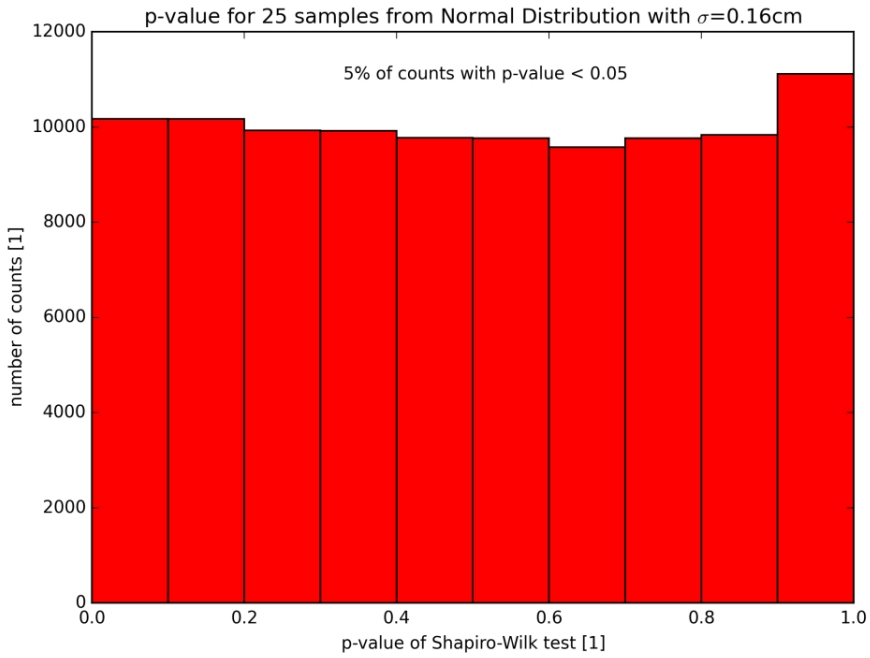


Figure 3.4: Histogram of p-values for setup error in head-feet direction. Simulation for 10^5 cases. Each case consist of 25 samples randomized from normal distribution of given standard deviation.

3.3.2 Time trends analysis

In order to check whether there is a time trend in setup errors for each patient and each direction, ordinary least square (OLS) fit was done. The two tailed p-values of the t-test for the parameters of fitted linear regression parameters were calculated as well. As the time trend was a potential source of not normal distributions of setup errors I have checked if the residuals (i.e. setup errors without time regression) would fulfill the normality assumption. This was tested again with Shapiro-Wilk test (see 3.3.1)

In order to see if there is a significant change in normality of pure setup errors and time trend Shapiro-Wilk statistic W achieved for both was compared and Wilcoxon signed-rank test was performed. The null hypothesis stated that there is no difference between Shapiro-Wilk statistic W values achieved for pure setup errors and for time regression residuals. Comparison of number of patients with normal error distribution in all directions for pure setup error and time regression residuals was also done.

The important question was whether existence of time trends could explain the setup errors distribution with low probability ($p < 0.05$) of having normal distribution. In order to answer that question a possible correlation between time regression slope and difference in p-value for pure and residual setup error was examined as well as the possible correlation between p-values of OLS fit and Shapiro-Wilk test.

3.3.3 Fitting probability distributions

Distributions of patient's specific characterisations of setup errors (like $\overline{m_p}$, SD_p) for Erasmus MC database were evaluated in order to find the probability distribution which would fit best to the data. Several functions were fitted to the distributions of mean ($\overline{m_p}$) and std (SD_p) of setup errors. The goodness of fit was calculated for each probability function as a residual sum of squares (RSS). The function with the smallest RSS for all directions was fitted and afterwards presented in comparison with the data histogram.

3.4 Results - COI

3.4.1 Testing normality

3.4.1.1 Data analysis for single patients

In Fig. 3.5 the worst and the best results of Shapiro-Wilk test for normality are shown. In left-right direction p-value varied between 0.0 and 0.894. In anterior-posterior direction p-value varied between 0.0 and 0.597. In head-feet direction p-value varied between 0.0 and 0.836.

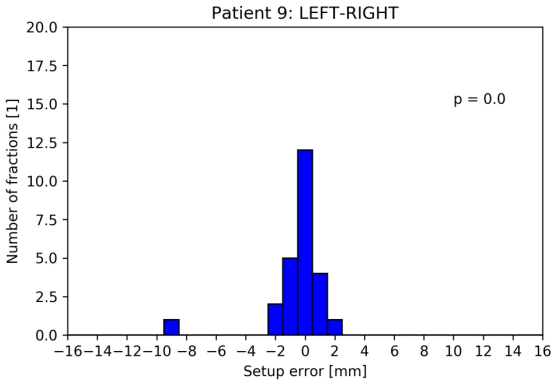
In Fig. 3.6-3.8 histograms of p-values achieved with Shapiro-Wilk test of setup errors for each patient are shown. In Fig. 3.9 p-values for all setup error directions are shown in order to make it easy to compare them. It can be seen that p-values distributions are similar. Best results in terms of normal distribution of setup errors were achieved for head-feet direction. If significance level is set to 0.05 the null hypothesis of normal distribution of setup errors can be rejected for 42% of patients in left-right and head-feet direction and for 43% of patients in anterior-posterior direction.

Setup errors have a spatial nature although all procedures (verification, margin calculations) are done per direction but in a combined manner. Due to that non-normal distribution in one direction can influence the whole patient procedure. Therefore the min p-value for each patient was also examined (see Fig. 3.10). With the confidence interval set to 95% the hypothesis of normal distribution of setup errors in all three directions was fulfilled in 22% of patients. All other patients had significant probability that distribution of their setup errors is not normal in at least one principal direction (see Fig. 3.11).

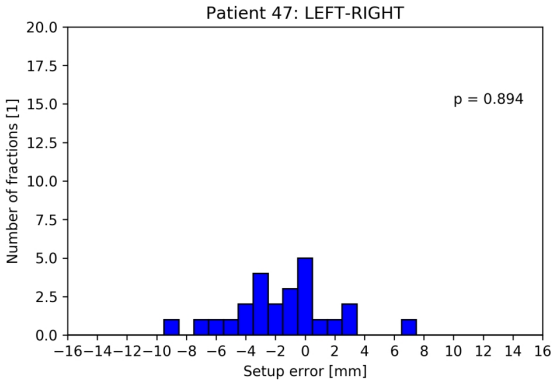
3.4.1.2 Data analysis for population of patients

Fig. 3.12 shows distribution of $\overline{m_p}$ which are interpreted as systematic errors. With the confidence interval set to 99%, the null hypothesis that systematic errors have normal distribution cannot be rejected for left-right and head-feet direction. The p-value for anterior-posterior direction is lower than 0.001. Despite the fact that not all of the patients have normal distribution of setup errors the population mean, M (see Eq. 2.5), of mean patient setup errors, $\overline{m_p}$ (see Eq. 2.3), was calculated. Spread of $\overline{m_p}$ distribution was also calculated as Σ (see Eq. 2.6). It was assumed that M was the best estimator of systematic error for whole population of patients. The values of Σ were: 1.99 mm for left-right, 2.98 mm for anterior-posterior and 1.65 mm for head-feet direction.

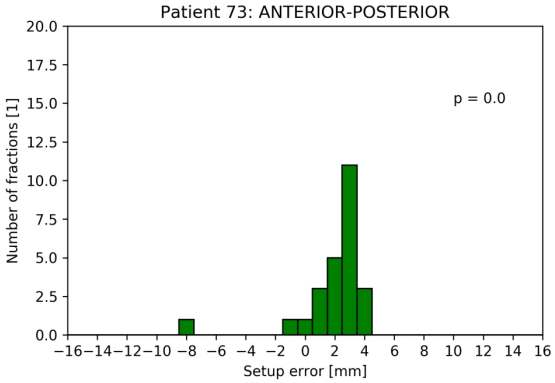
Variability of the SD_p (see Eq. 2.4) which was assumed to be the best estimator of the standard deviation of patient's random errors was also calculated. As it is shown in Fig. 3.12 there is significant difference between the SD_p distribution and normal distribution for all



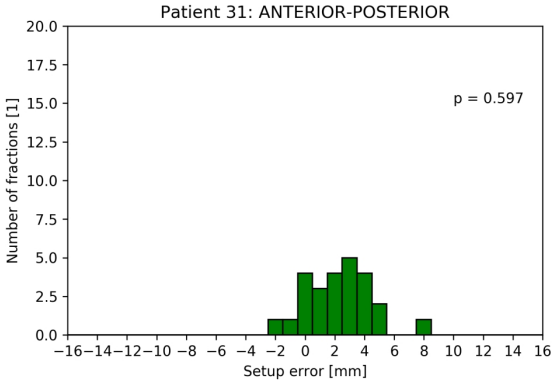
(a) Patient with the lowest observed p-value



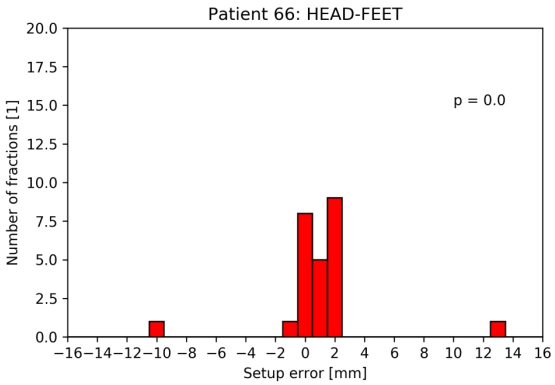
(b) Patient with the highest observed p-value



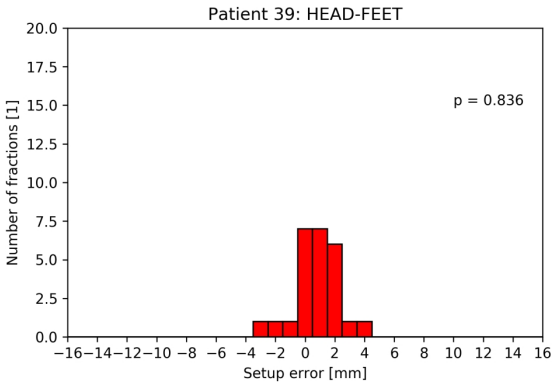
(c) Patient with the lowest observed p-value



(d) Patient with the highest observed p-value



(e) Patient with the lowest observed p-value



(f) Patient with the highest observed p-value

Figure 3.5: Histogram of setup errors distributions in left-right direction (3.5a and 3.5b), anterior-posterior direction (3.5c and 3.5d) and head-feet direction (3.5e and 3.5f). In the left column distribution of setup errors with the lowest p-value is presented. In the right column distribution of setup errors with the highest p-value is presented. Data achieved for COI database.

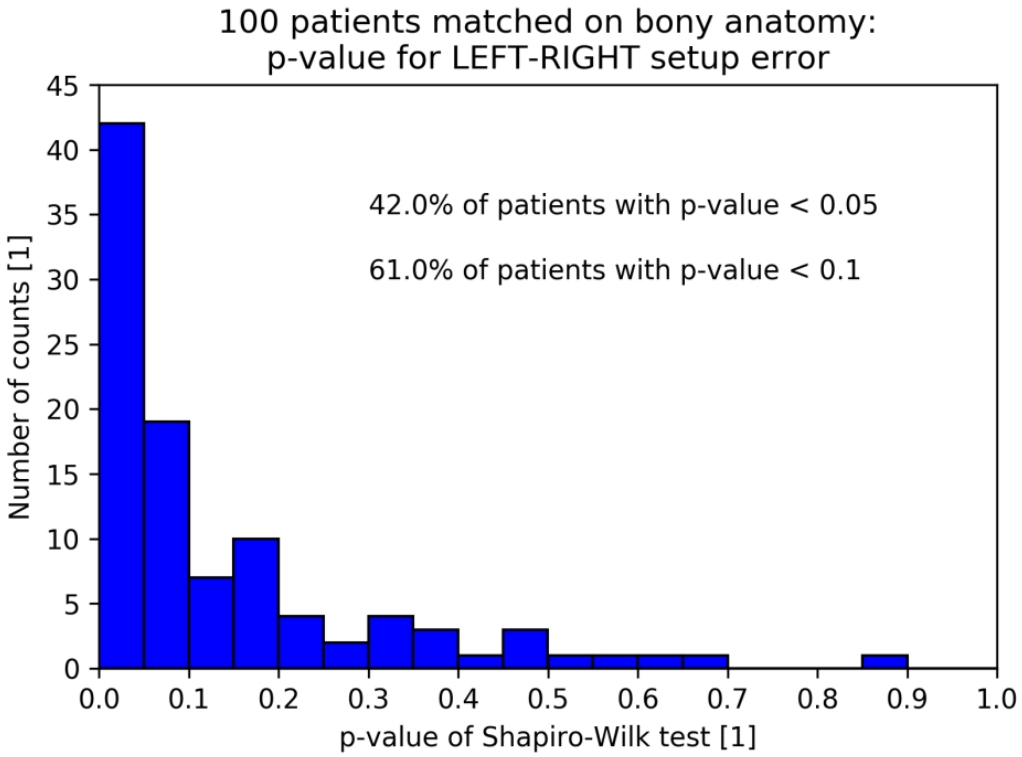


Figure 3.6: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in left-right direction. Data achieved for COI database.

principal directions. Values of σ (see Eq. 2.7) were: 2.23 mm for left-right, 2.05 mm for anterior-posterior and 1.53 mm for head-feet direction. The width of SD_p distribution, denoted as SD_{SD}, was also calculated and was equal: 0.64 mm, 0.65 mm and 0.43 mm for left-right, anterior-posterior and head-feet direction respectively — see Fig. 3.12.

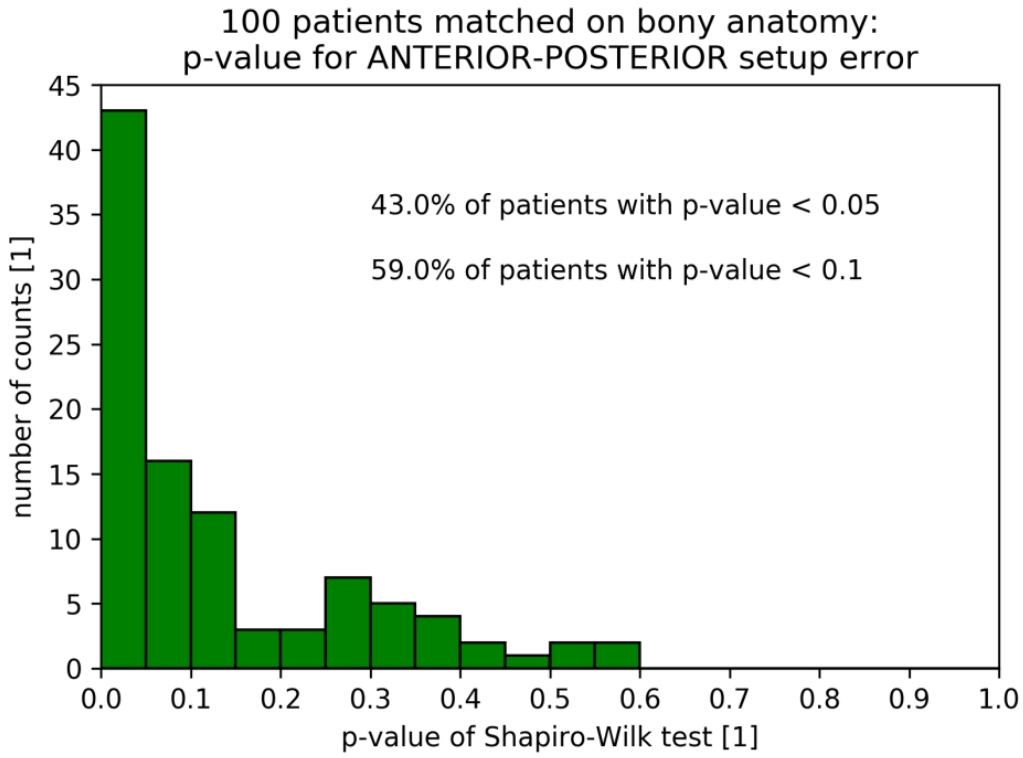


Figure 3.7: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in anterior-posterior direction. Data achieved for COI database.

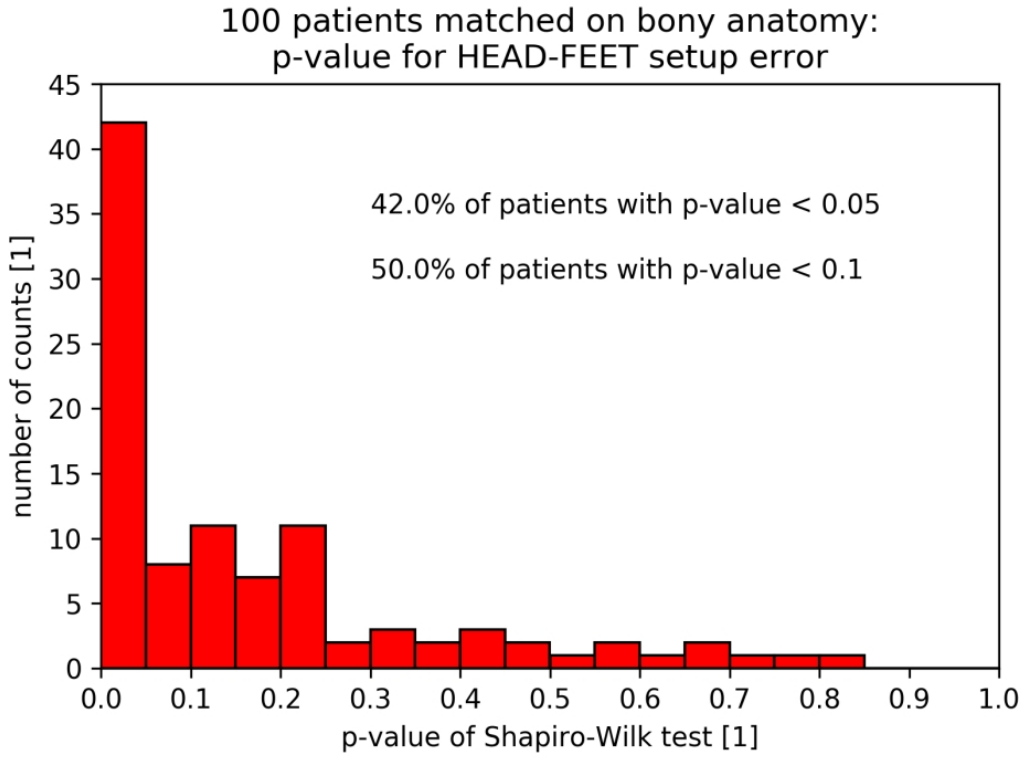


Figure 3.8: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in head-feet direction. Data achieved for COI database.

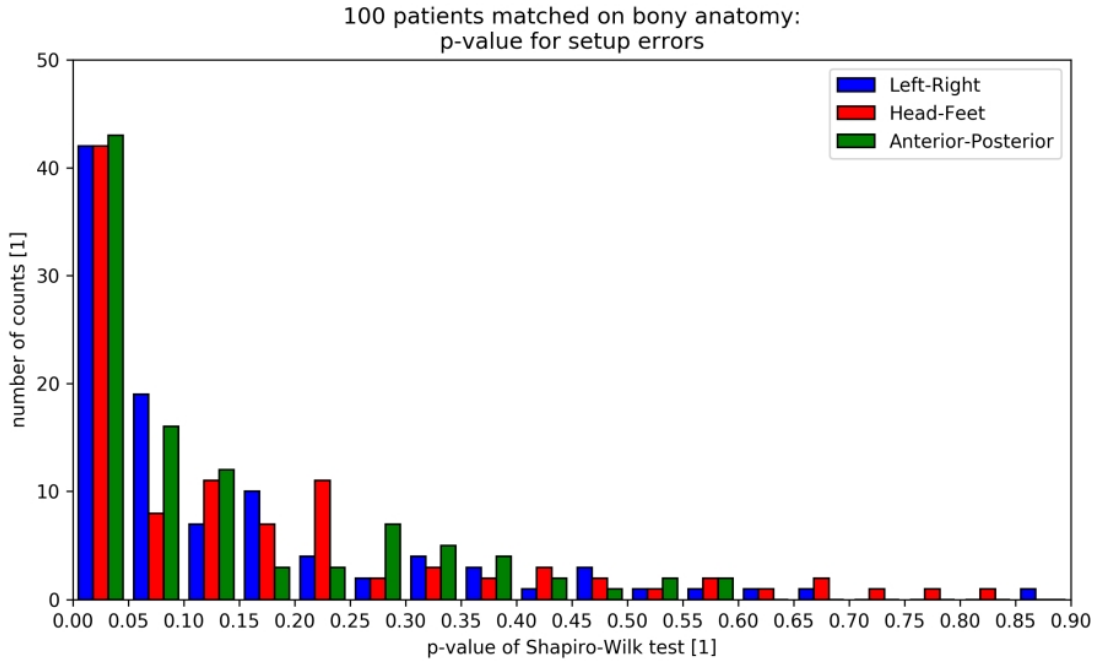


Figure 3.9: Comparison of histograms of p-values achieved with Shapiro-Wilk test for setup errors in all directions. Data achieved for COI database.

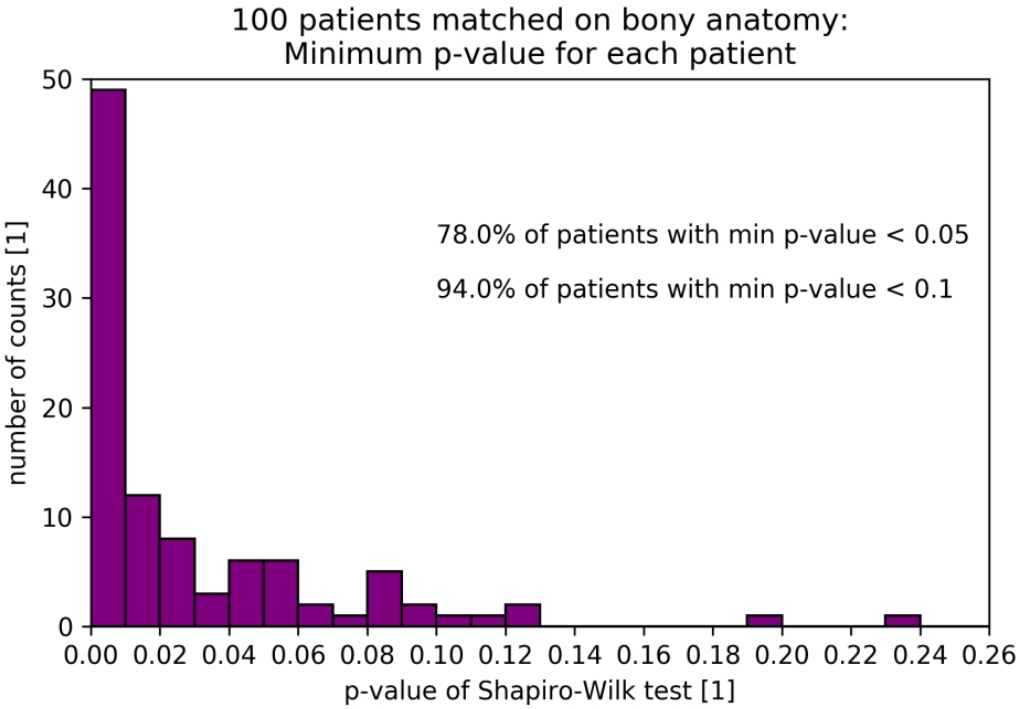


Figure 3.10: Histogram of minimum p-value, out of three directions, for all patients. Data achieved for COI database.

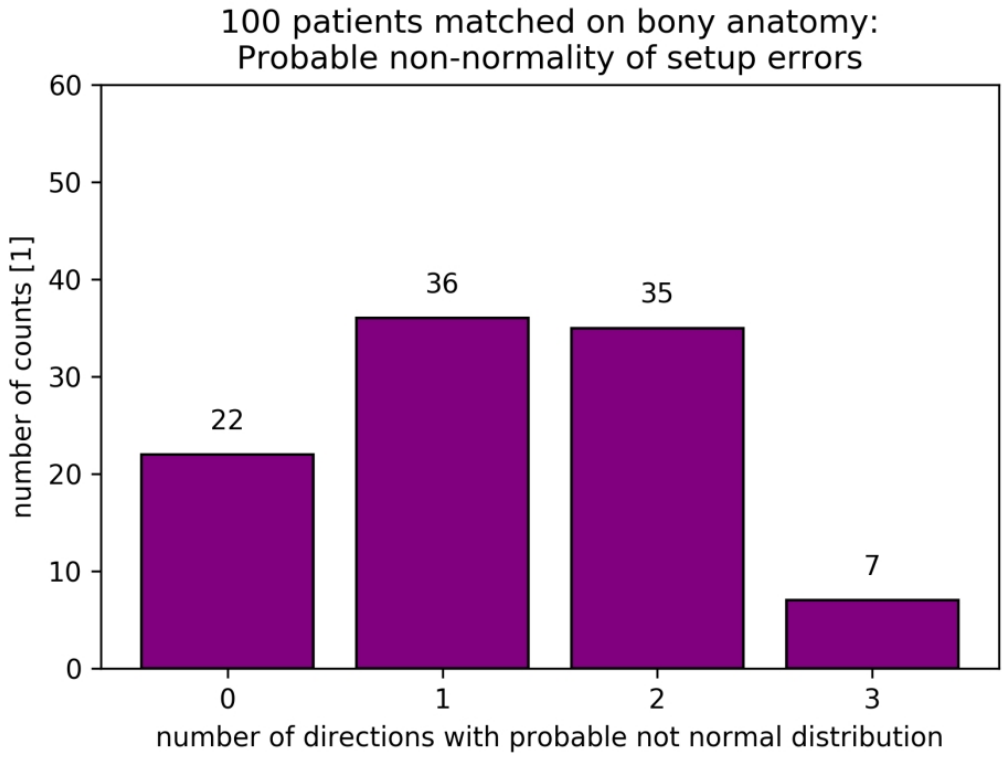
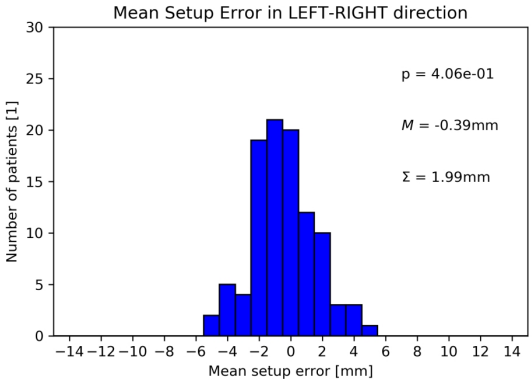
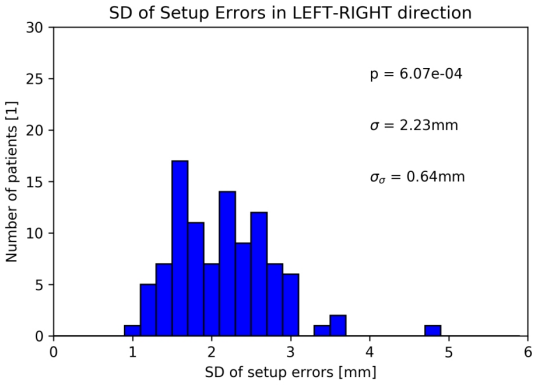


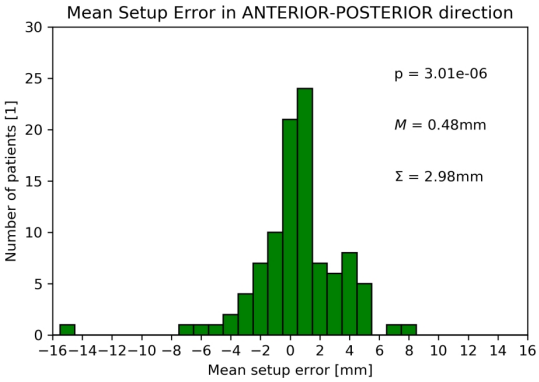
Figure 3.11: Number of patients with probable not normal setup error distribution in 0-1-2-3 directions. Confidence interval was set to 95%. Data achieved for COI database.



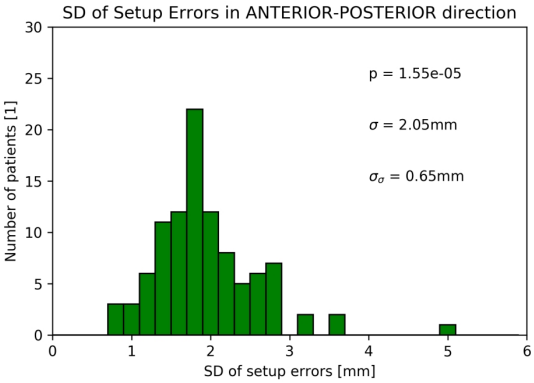
(a) Distribution of $\overline{m_p}$



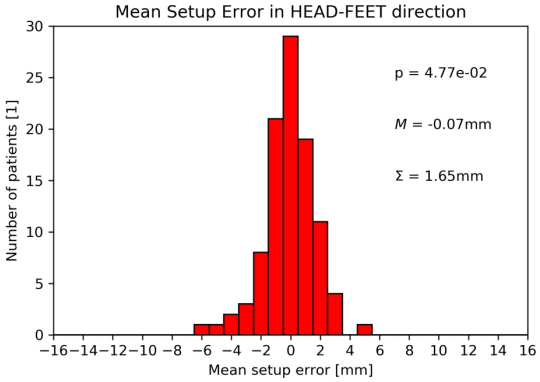
(b) Distribution of SD_p



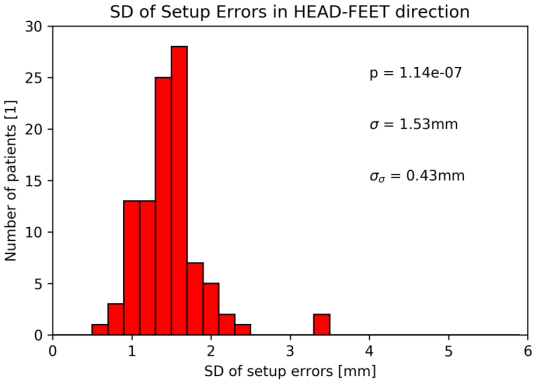
(c) Distribution of $\overline{m_p}$



(d) Distribution of SD_p



(e) Distribution of $\overline{m_p}$



(f) Distribution of SD_p

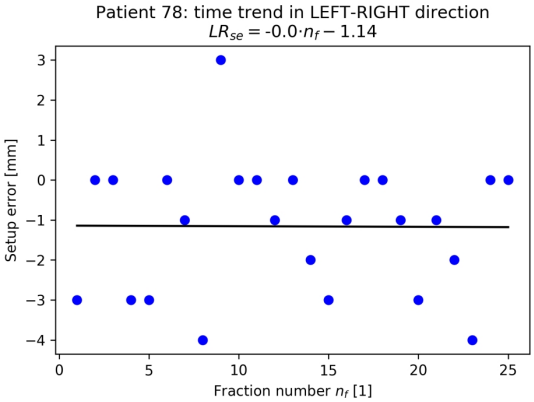
Figure 3.12: Histograms of $\overline{m_p}$ and SD_p for setup errors in left-right direction (3.12a and 3.12b), anterior-posterior direction (3.12c and 3.12d) and head-feet direction (3.12e and 3.12f). The p-value of Shapiro-Wilk test is written on each histogram. Data achieved for COI database.

3.4.2 Time trends analysis

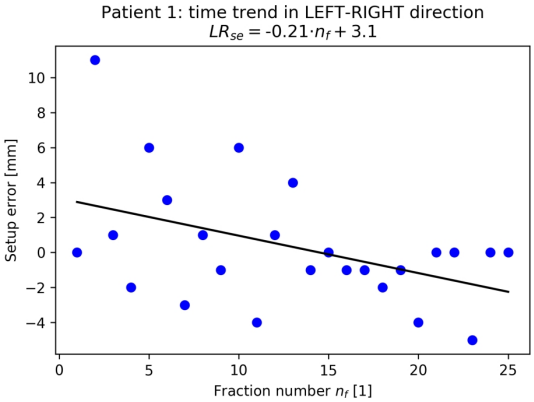
Absolute value of slope for time regression fit varied from 0.0 to 0.21 mm/fraction for left-right direction, from 0.0 to 0.47 mm/fraction for anterior-posterior direction, and from 0.0 to 0.14 mm/fraction for head-feet direction (see Fig. 3.13). The biggest observed negative change in time was about -0.5 mm/fraction. That leads to change in patient position of 5 mm per 10 fractions. In that situation setup error is a decreasing function of time. The greatest observed positive change in time was about 0.3 mm/fraction. In that situation setup error is an increasing function of time. It has to be noticed that not only slope value has an impact on behaviour of setup errors during treatment but also offset (initial setup error). In order to check if the absolute setup error value decreases or increases during the course of treatment the difference between the absolute setup error in last and first fraction has to be examined. That has to be done not for random errors but for errors lying on the fitted time trend line (denoted as trendline errors in this thesis). The concept of dividing setup errors for random errors and trendline errors was depicted and widened in Section 4.2.2.2 and Section 7.2.1 Fig. 3.15-3.17 show that for all principal directions slightly more patients experienced increase in setup error during course of treatment. Histograms of slope values are shown in Fig. 3.14.

Wilcoxon signed-rank test showed that there is a significant difference between Shapiro-Wilk W statistic values obtained for pure setup errors and for that obtained for time regression residuals – p-values were of order 10^{-14}-10^{-15} (see Fig. 3.18-3.20). Despite that normal distribution still cannot be assumed for about 20% of patients if the significance level is set to 5% (22% for left-right, 18% for anterior-posterior and 20% for head-feet direction). Due to this it can be concluded that non-normality problem can be explained by the existence of time trend only in half of the cases. Detailed results are shown in Fig. 3.21-3.23.

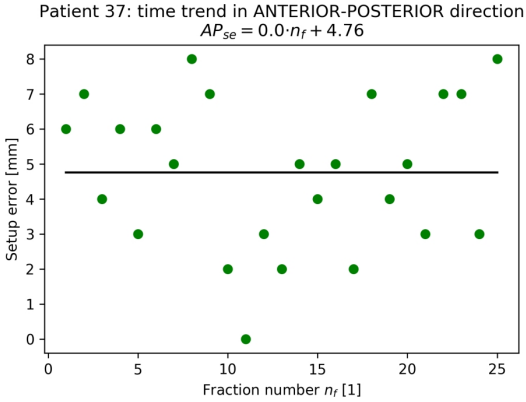
Fig. 3.24 shows that there is a significant difference between number of patients with normal distribution in setup errors in all directions if we consider time regression residuals. Despite that for 48% of patients distribution of setup error is rather not normal for at least one direction. It has to be emphasized that for some patients the Shapiro-Wilk test p-value was lower for residual setup errors than for pure ones. Some setup errors distributions which appeared to be normal for pure errors seemed to be not normal for time trend residuals.



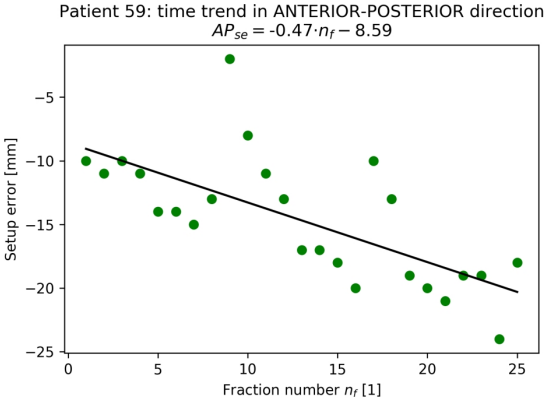
(a) Patient with the lowest observed slope



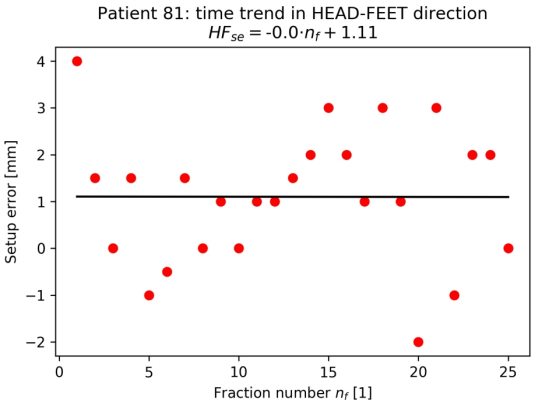
(b) Patient with the highest observed slope



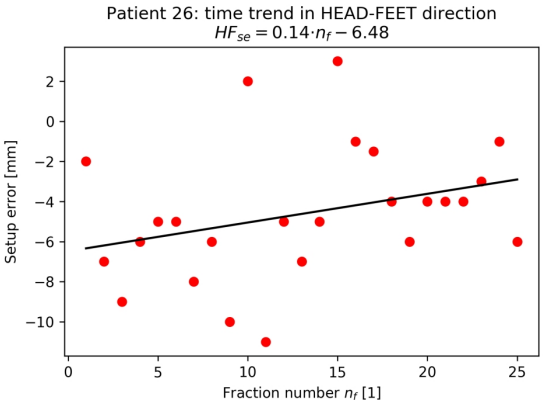
(c) Patient with the lowest observed slope



(d) Patient with the highest observed slope

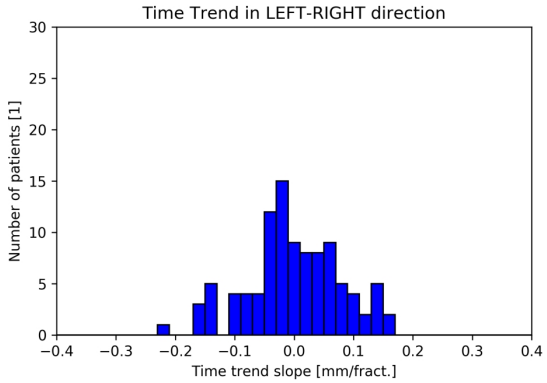


(e) Patient with the lowest observed slope

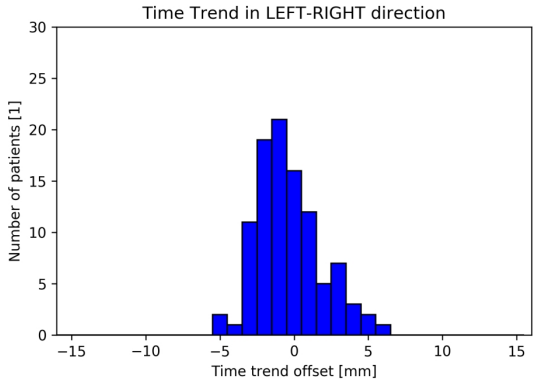


(f) Patient with the highest observed slope

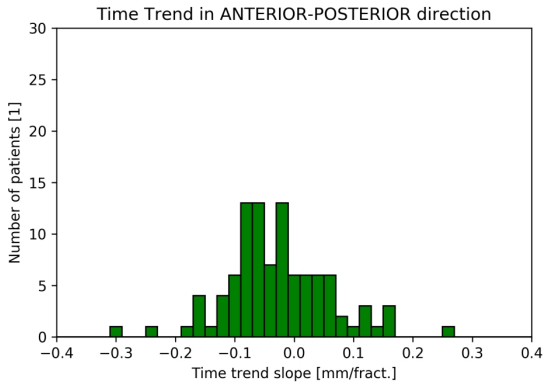
Figure 3.13: Plots of setup error distribution along left-right axis (3.13a and 3.13b), anterior-posterior axis (3.13c and 3.13d) and head-feet axis (3.13e and 3.13f). On the left patients with the lowest observed absolute slope values are shown (i.e. patients with no observed time trend), on the right patients with highest observed slope values are shown (significant time trend observed). LR_{se}, AP_{se}, HF_{se} stands for m_{pf} in particular direction while n_f describes fraction number and is a measure of time. Data achieved for COI database.



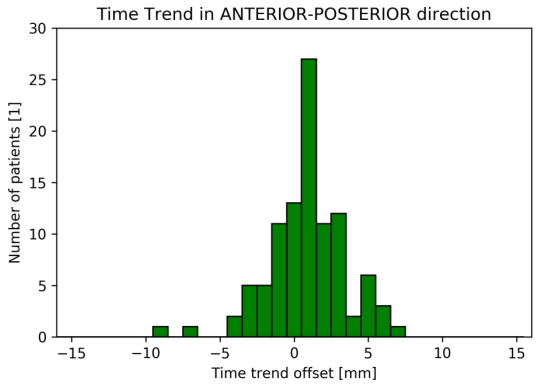
(a) Histogram of slope values



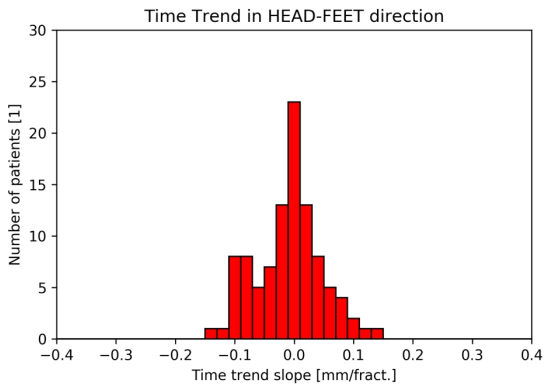
(b) Histogram of offset values



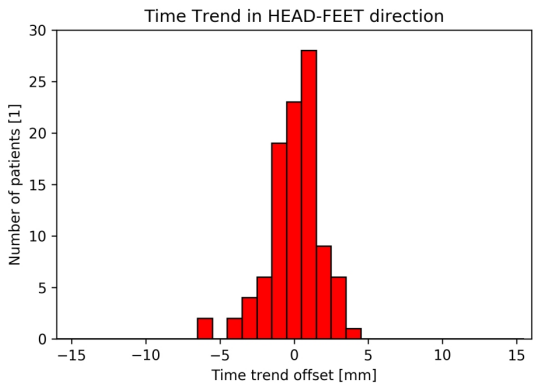
(c) Histogram of slope values



(d) Histogram of offset values



(e) Histogram of slope values



(f) Histogram of offset values

Figure 3.14: Histograms of slope (left) and offset (right) values for time trend OLS analysis of setup errors in each direction. Data achieved for COI database.

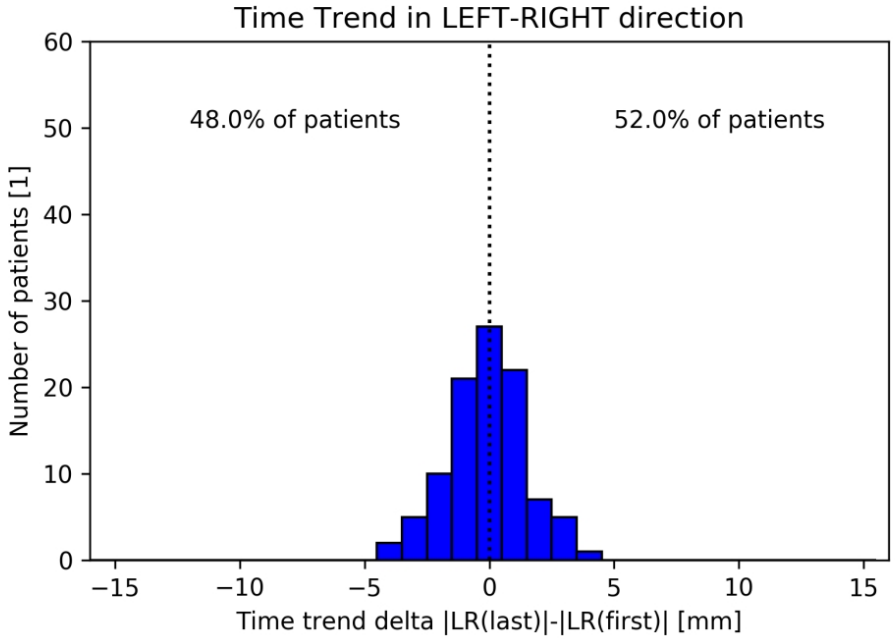


Figure 3.15: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database.

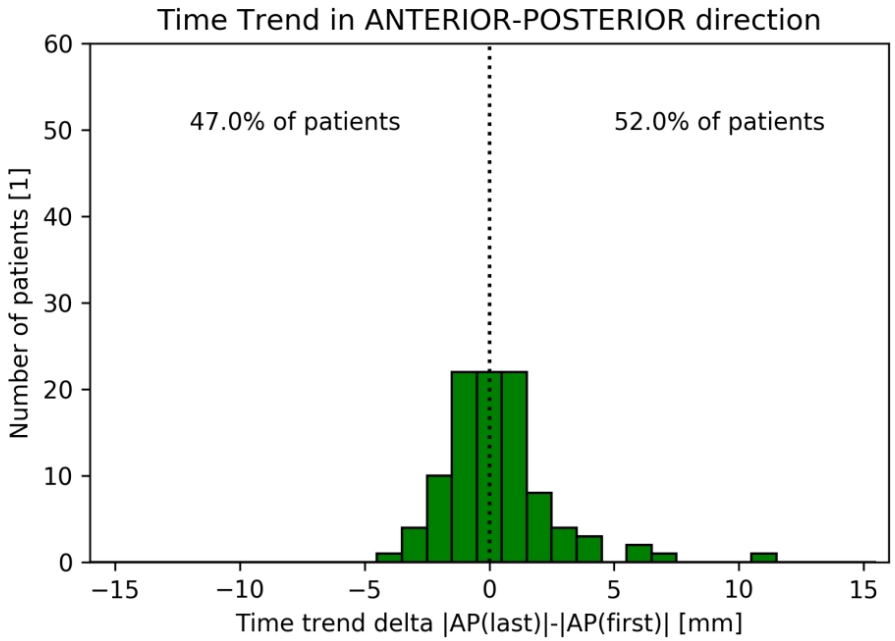


Figure 3.16: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database.

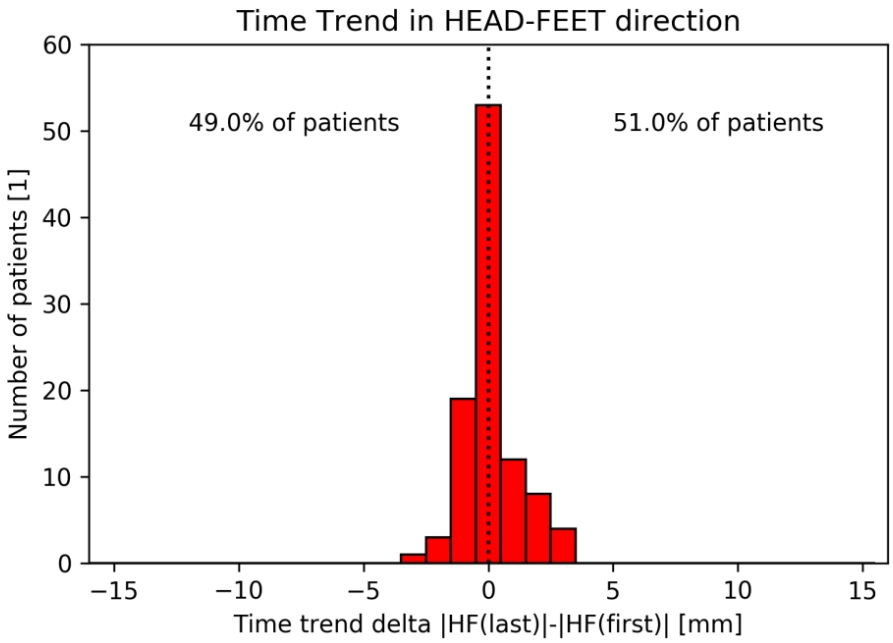


Figure 3.17: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for COI database.

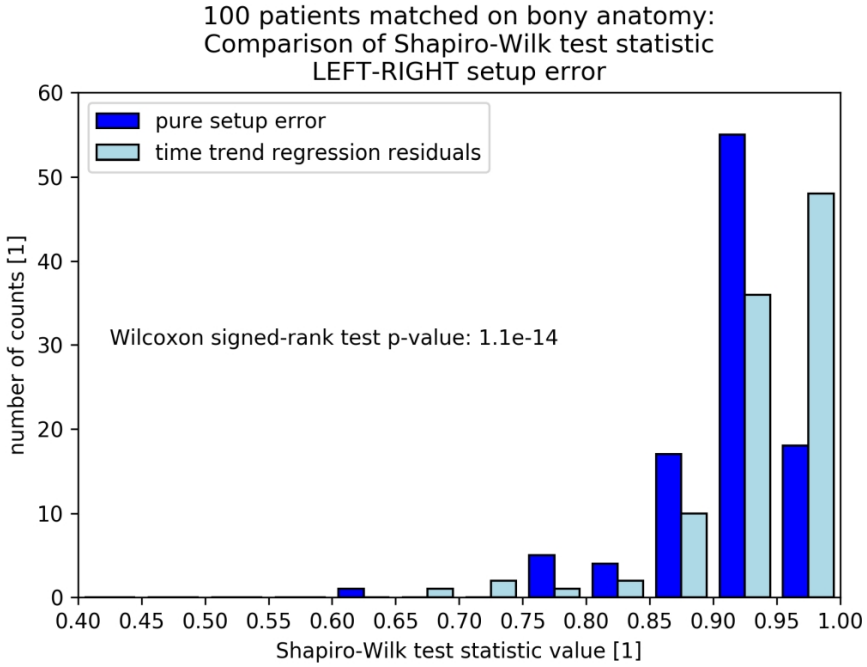


Figure 3.18: Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in left-right direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database.

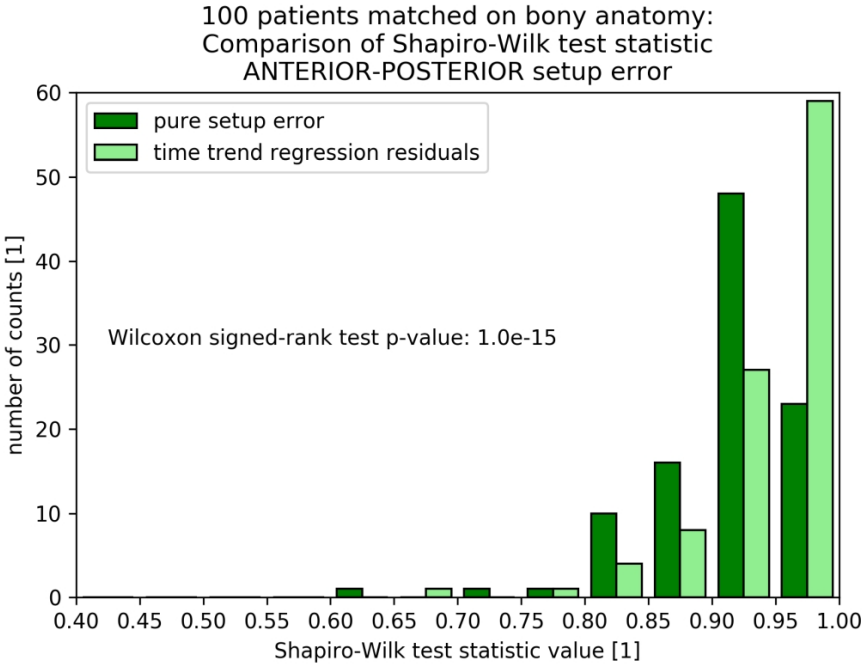


Figure 3.19: Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in anterior-posterior direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database.

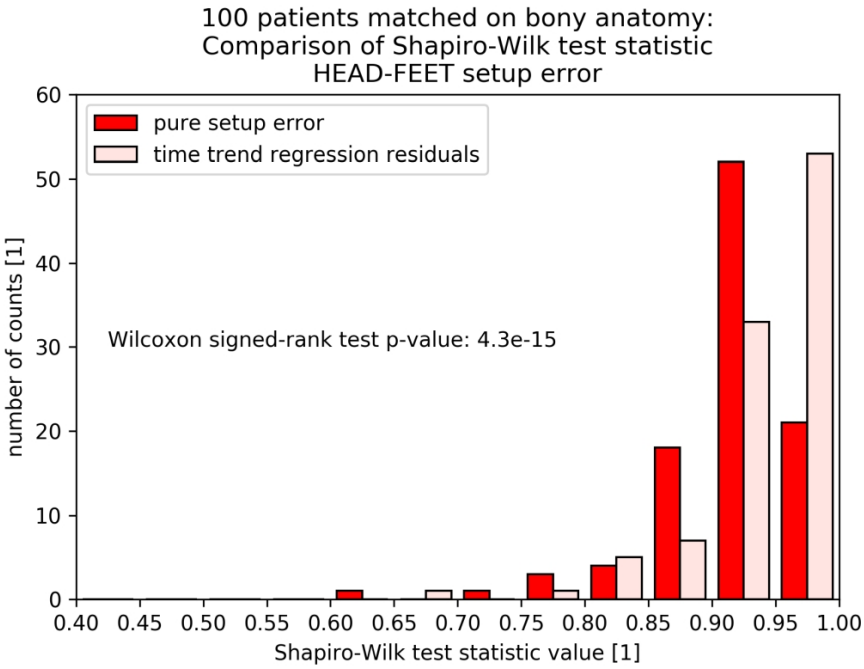


Figure 3.20: Histograms of Shapiro-Wilk W statistic values achieved during normality test of setup errors in head-feet direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for COI database.

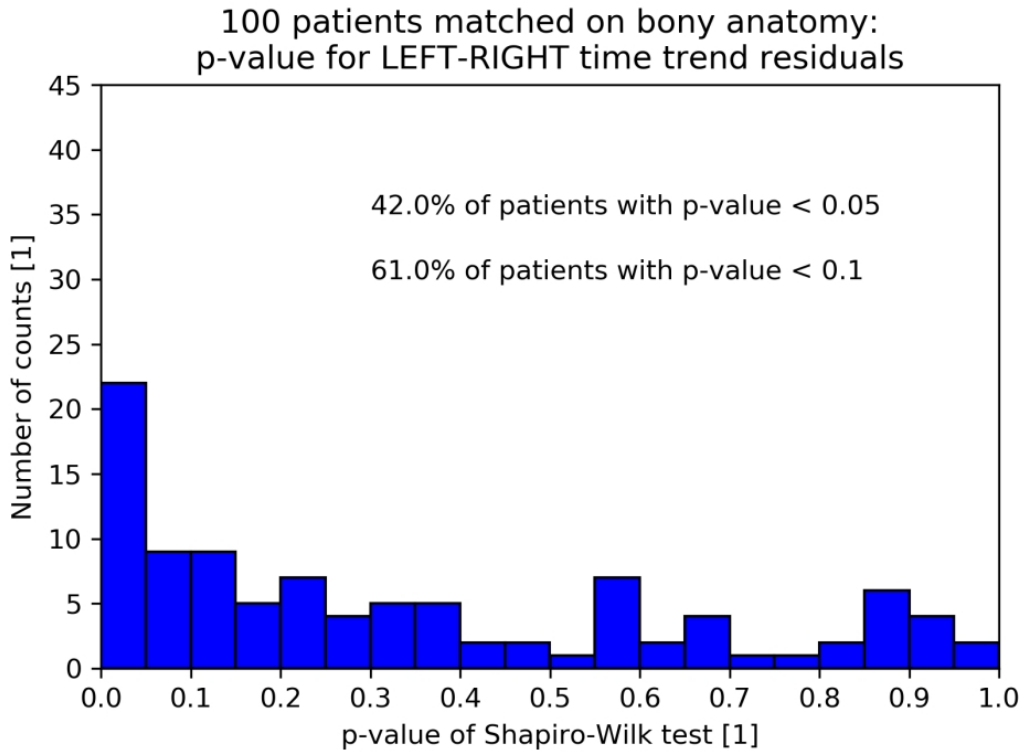


Figure 3.21: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Left-Right direction. Data achieved for COI database.

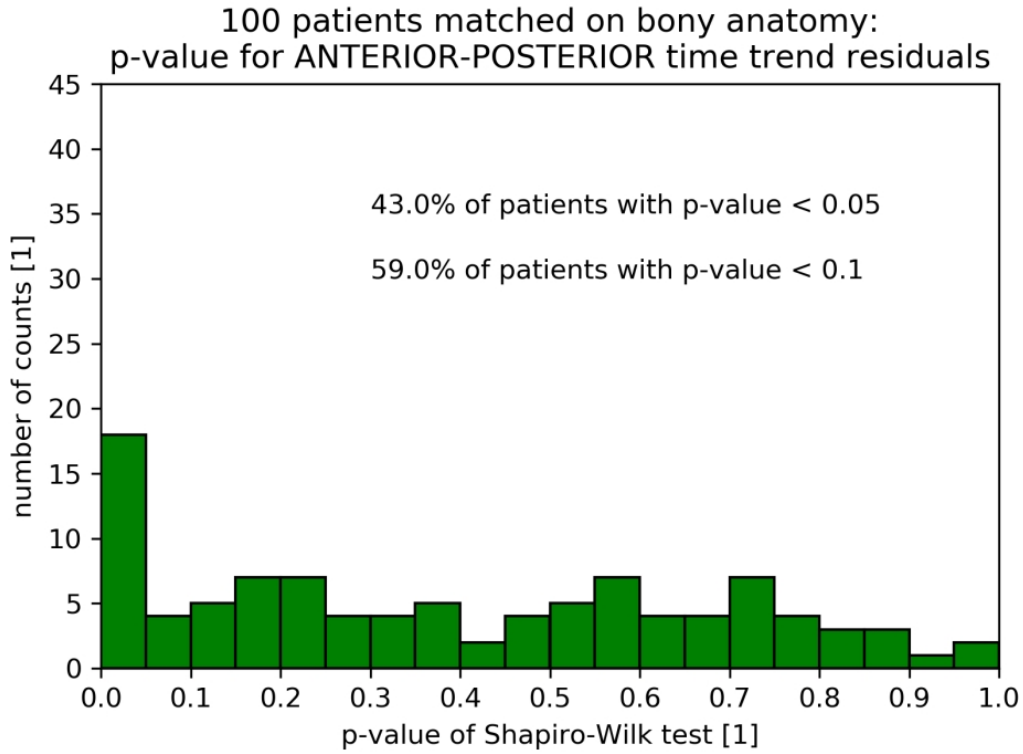


Figure 3.22: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Anterior-Posterior direction. Data achieved for COI database.

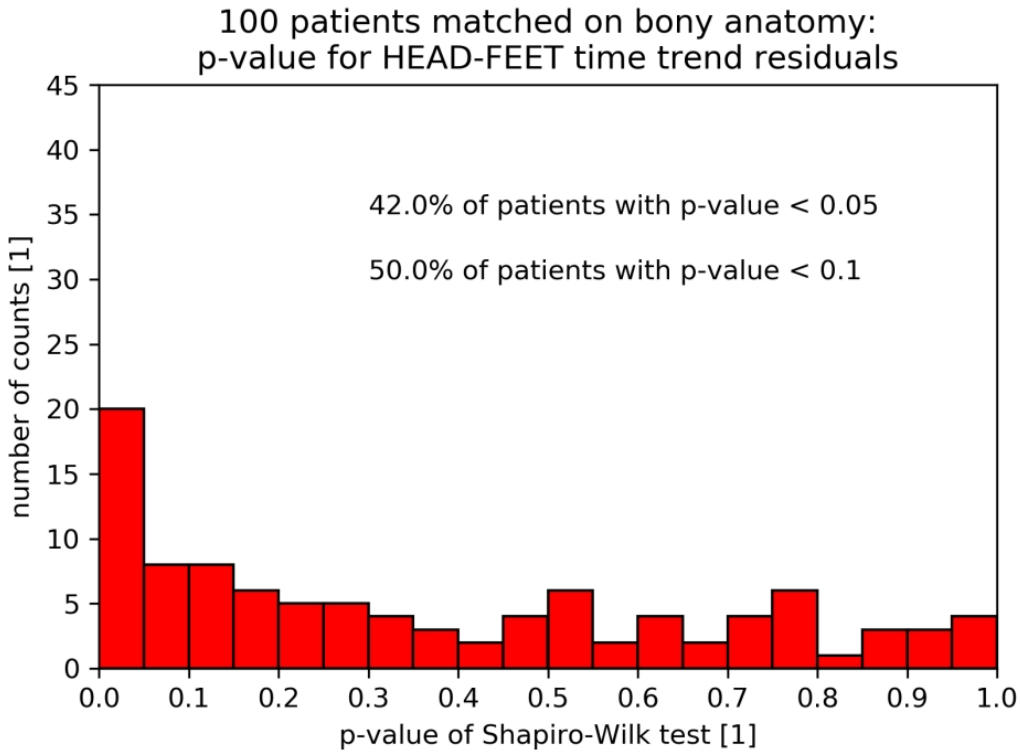


Figure 3.23: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in Head-Feet direction. Data achieved for COI database.

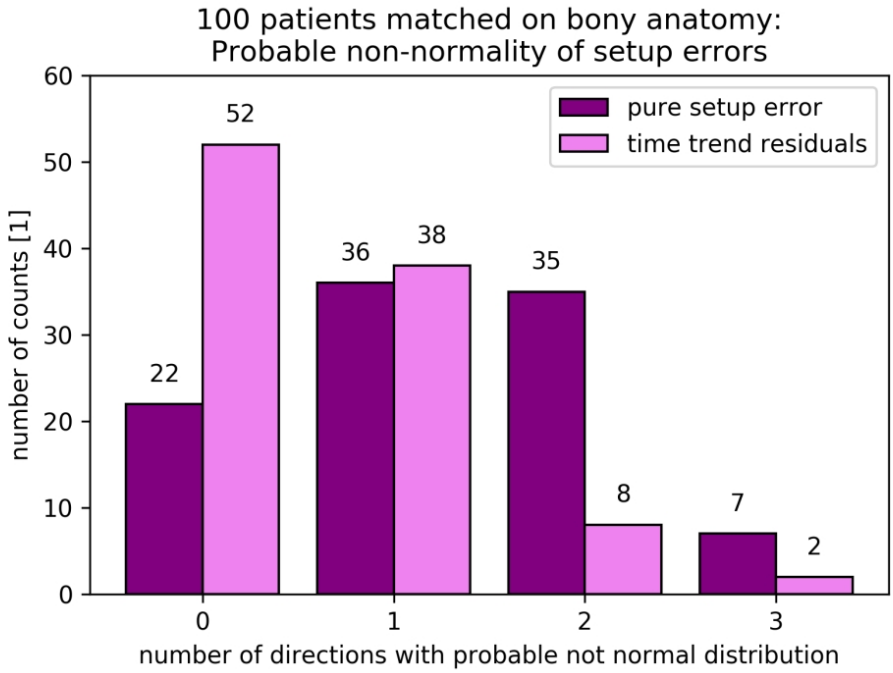


Figure 3.24: Histograms of number of patients with specified number of directions with probable not normal distribution — comparison between pure setup errors and time trend residuals. Data achieved for COI database.

3.5 Results - Erasmus MC

3.5.1 Testing normality

3.5.1.1 Data analysis for single patients

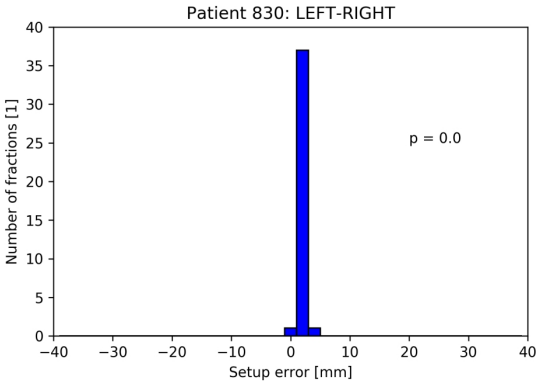
In Fig. 3.25, for each principal direction, setup errors distributions with the worst and the best p-value of Shapiro-Wilk test for normality are shown. In all directions these p-values varied between 0.0 and, approximately, 1.0.

In Fig. 3.26-3.28 histograms of p-values achieved with Shapiro-Wilk test of setup errors for each patient are shown. Distributions of p-values for all principal directions are similar (see Fig. 3.29). Best results in terms of normal distribution of setup errors were achieved for head-feet direction. If significance level is set to 0.05 the null hypothesis of normal distribution of setup errors can be rejected for 12.1% of patients in left-right, for 15.7% patients in anterior-posterior direction and for 10.5% of patients in head-feet direction. With the change of significance level to 0.1 these values were 18.4%, 22.5% and 17.2% respectively.

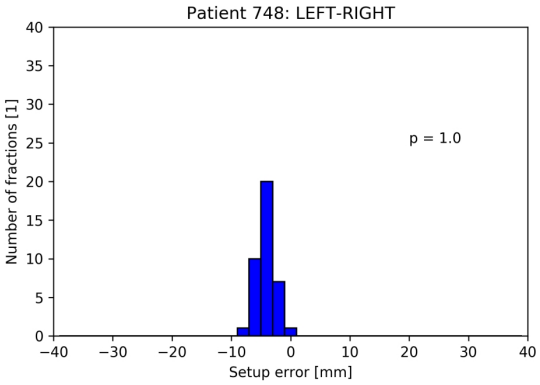
Similarly to the COI database the min p-value for each patient other three principal directions was also examined (see Fig. 3.30). With the confidence interval set to 95% the hypothesis of normal distribution of setup errors in all three directions was fulfilled in 67.3% of patients. The hypothesis of normal distribution of setup errors in at least two directions was fulfilled in 94.7% of patients (see Fig. 3.31).

3.5.1.2 Data analysis for population of patients

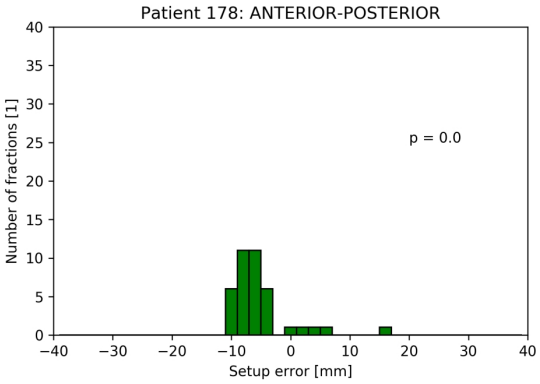
The mean value of setup error for each patient, $\overline{m_p}$, was calculated and assumed to be the best estimator of systematic error. The standard deviation of setup errors for each patient — SD_p was also calculated. As it is shown in Fig. 3.32 in none of the three directions $\overline{m_p}$ nor SD_p have the normal distribution (see Fig. 3.32). The values of Σ, i.e. the width of $\overline{m_p}$ distribution, were: 2.50 mm for left-right, 3.47 mm for anterior-posterior and 3.37 mm for head-feet direction. Analogically the values of σ (which is the RMS of SD_p) were: 1.93 mm for left-right, 2.77 mm for anterior-posterior and 2.64mm for head-feet direction. The width of SD_p distributions, SD_{SD}, was equal 0.71 mm, 0.80 mm and 0.68 mm respectively — see Fig. 3.32.



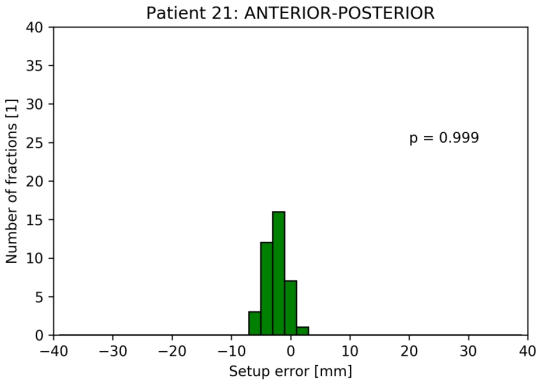
(a) Patient with the lowest observed p-value



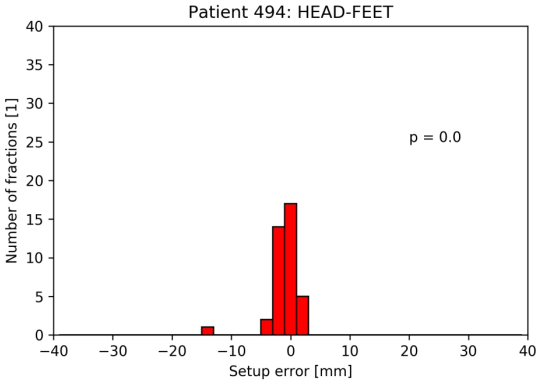
(b) Patient with the highest observed p-value



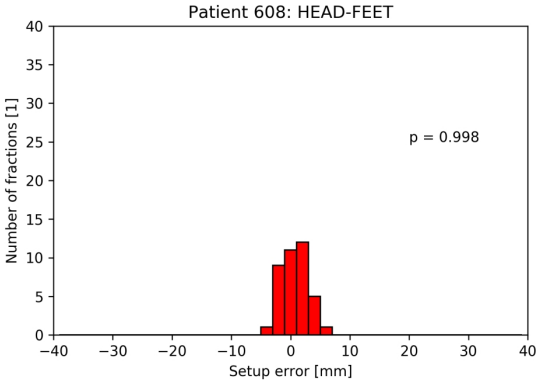
(c) Patient with the lowest observed p-value



(d) Patient with the highest observed p-value



(e) Patient with the lowest observed p-value



(f) Patient with the highest observed p-value

Figure 3.25: Histograms of setup error distribution in left-right direction (3.25a and 3.25b), anterior-posterior direction (3.25c and 3.25d) and head-feet direction (3.25e and 3.25f). Data achieved for Erasmus MC database.

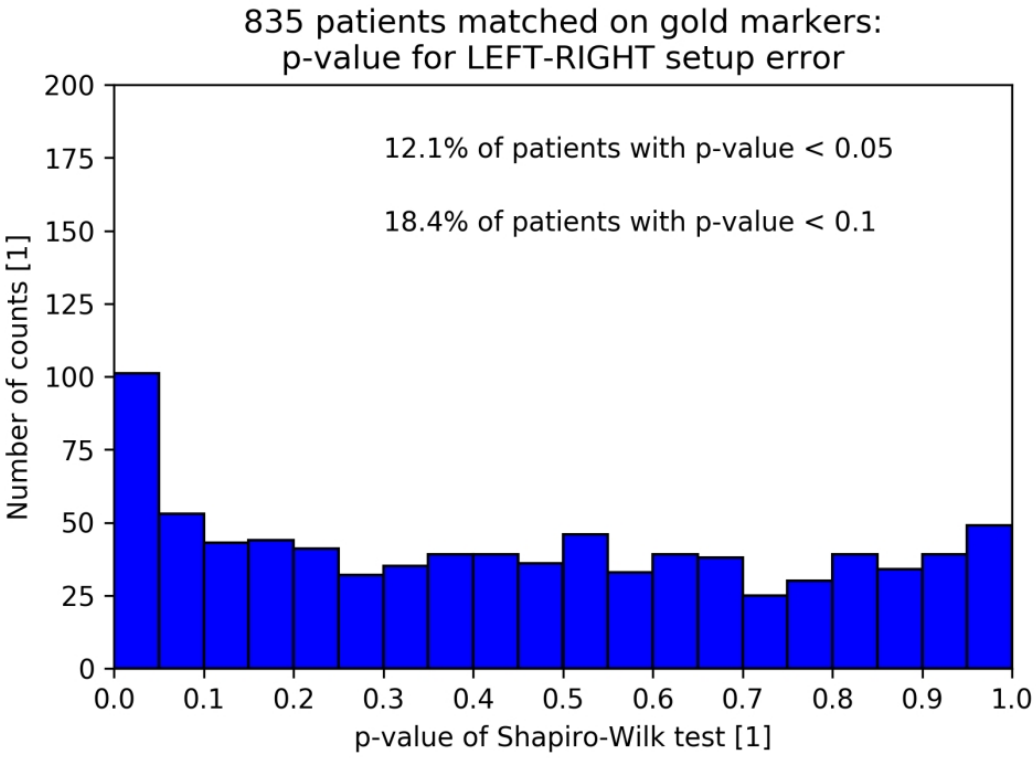


Figure 3.26: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in left-right direction. Data achieved for Erasmus MC database.

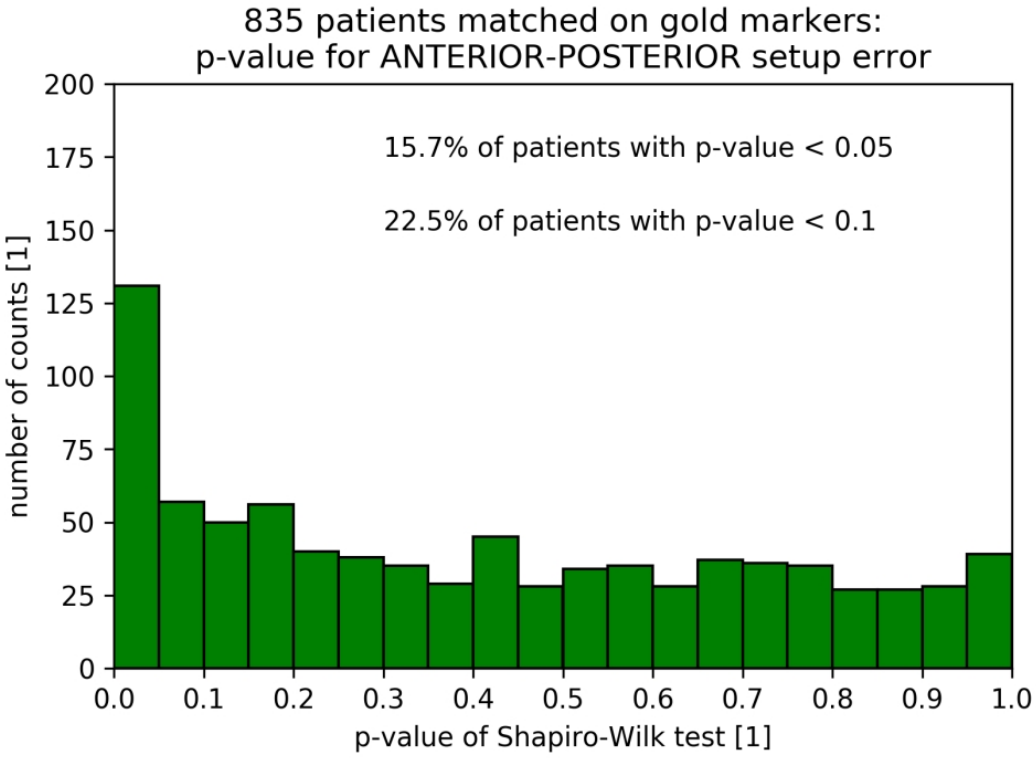


Figure 3.27: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in anterior-posterior direction. Data achieved for Erasmus MC database.

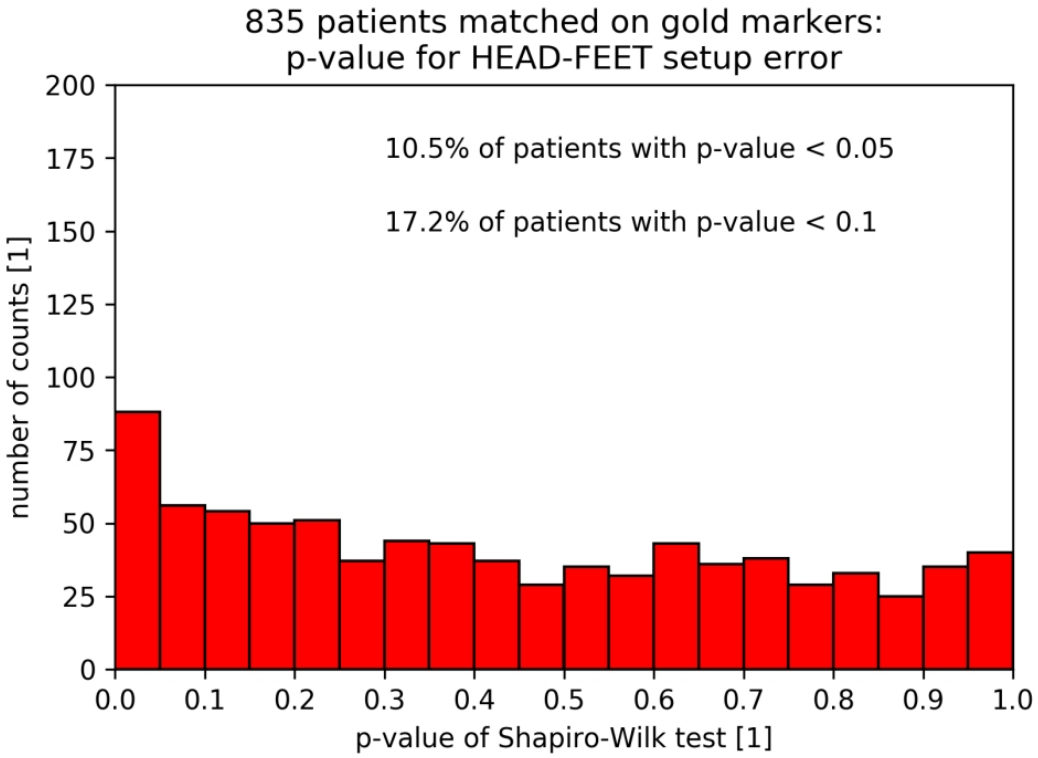


Figure 3.28: Histogram of p-values achieved with Shapiro-Wilk test for setup errors in head-feet direction. Data achieved for Erasmus MC database.

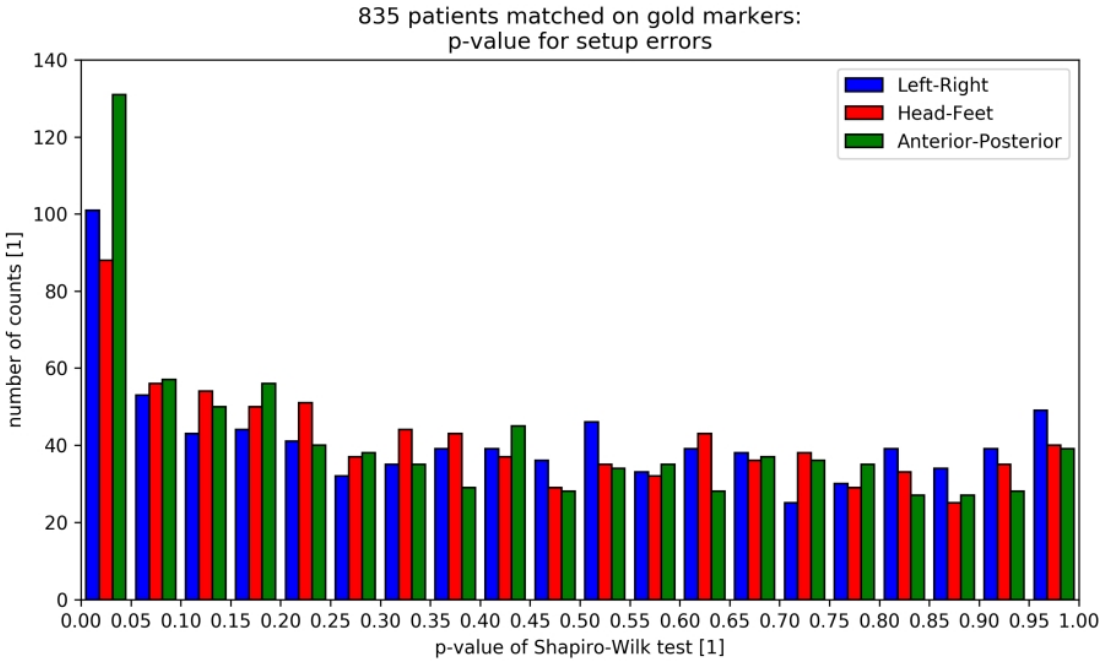


Figure 3.29: Comparison of histograms of p-values achieved with Shapiro-Wilk test for setup errors in all directions. Data achieved for Erasmus MC database.

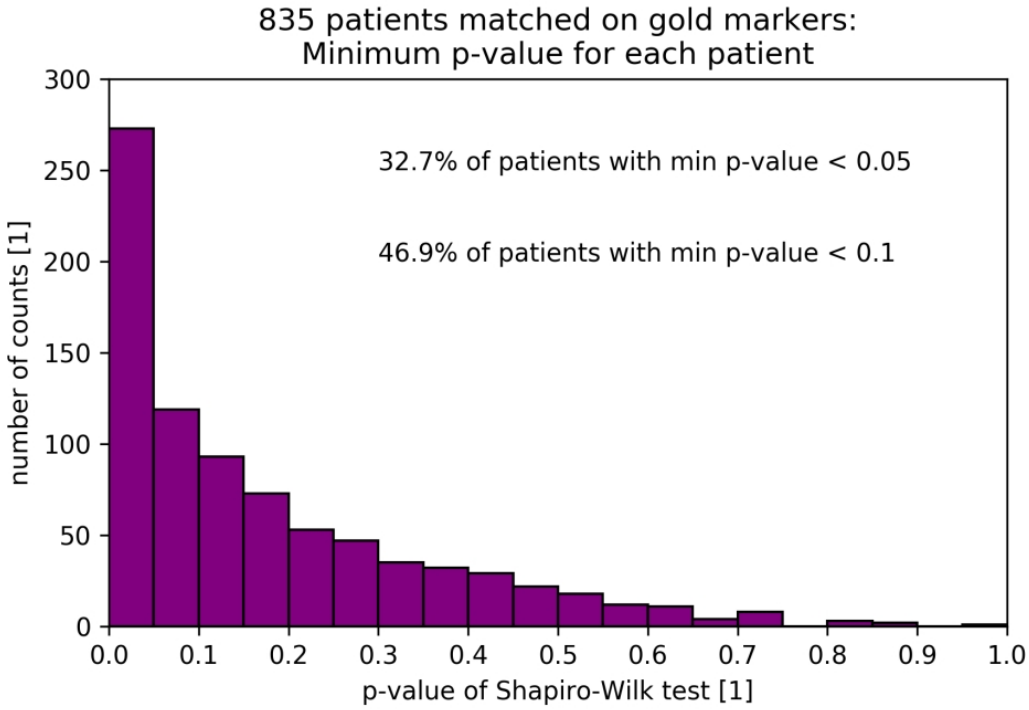


Figure 3.30: Histogram of minimum p value, out of three directions, for all patients. Data achieved for Erasmus MC database.

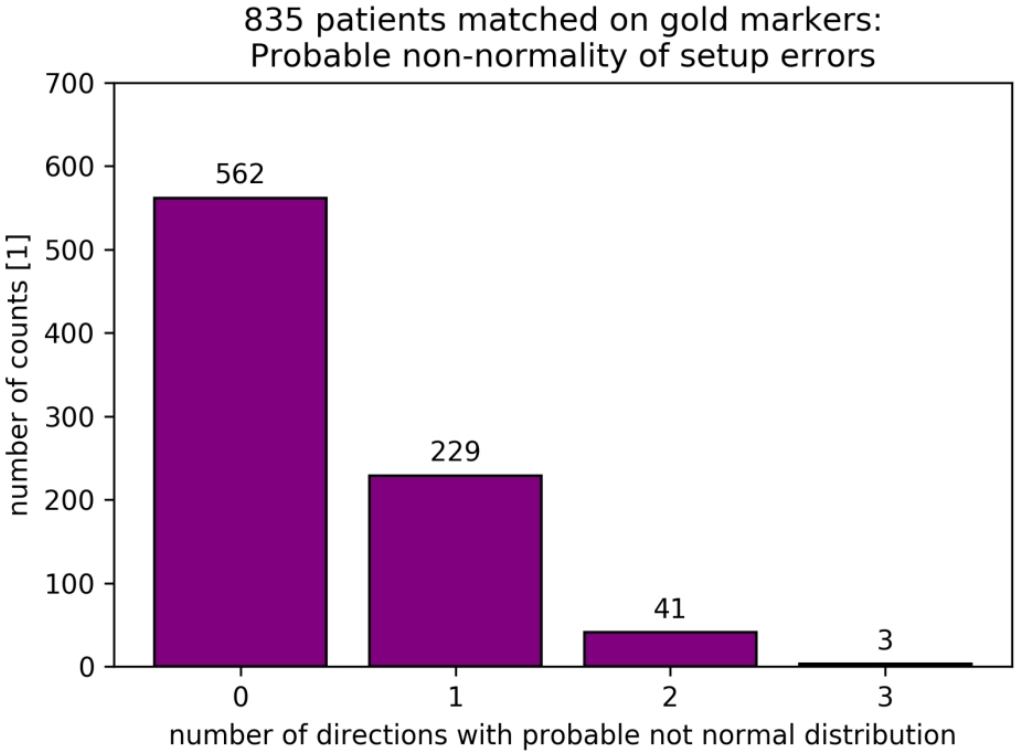
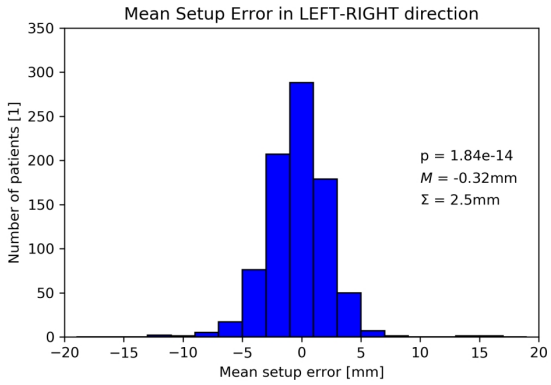
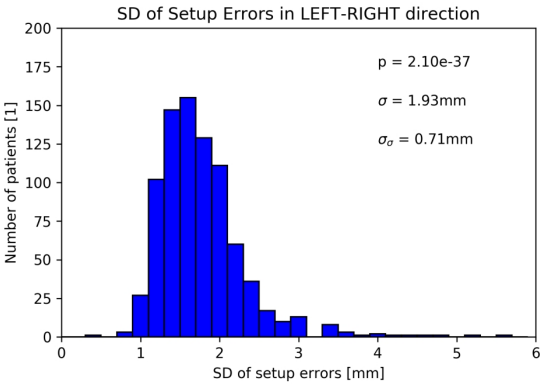


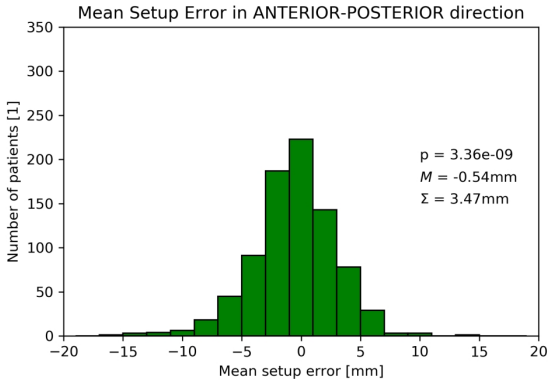
Figure 3.31: Number of patients with probable not normal setup error distribution in 0-1-2-3 directions. Confidence interval was set to 95%. Data achieved for Erasmus MC database.



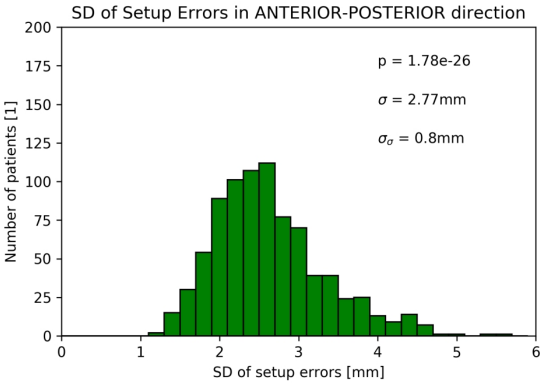
(a) Distribution of $\overline{m_p}$



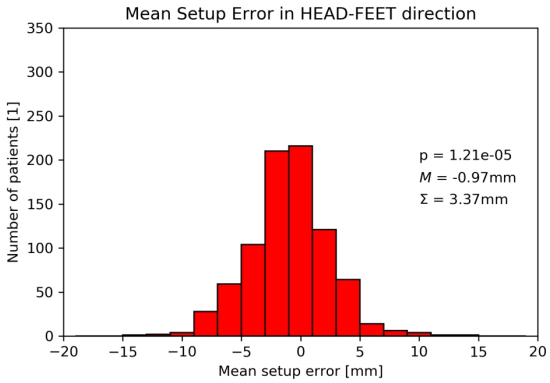
(b) Distribution of SD_p



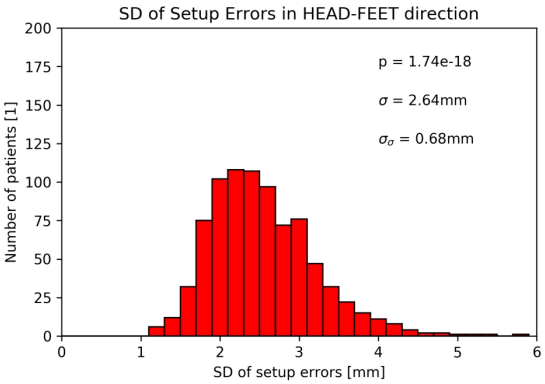
(c) Distribution of $\overline{m_p}$



(d) Distribution of SD_p



(e) Distribution of $\overline{m_p}$



(f) Distribution of SD_p

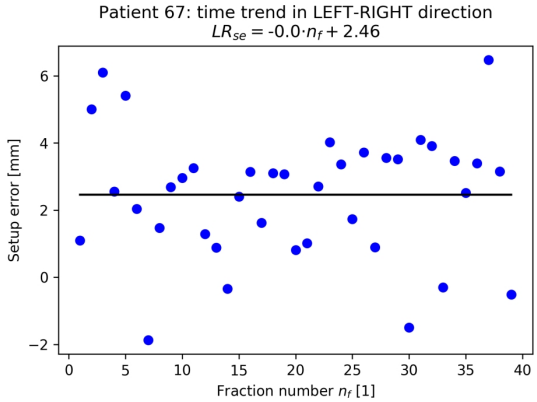
Figure 3.32: Histograms of $\overline{m_p}$ and SD_p for setup errors in left-right direction (3.32a and 3.32b), anterior-posterior direction (3.32c and 3.32d) and head-feet direction (3.32e and 3.32f). The p-value of Shapiro-Wilk test comparing normal ditribution with presented histogram is written on each figure. Data achieved for Erasmus MC database.

3.5.2 Time trends analysis

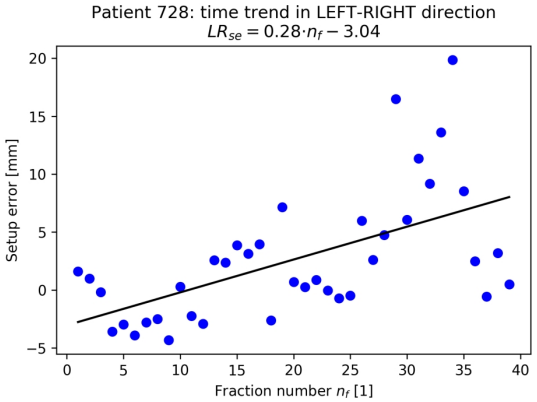
Absolute value of slope for time regression fit varied from 0.0 to 0.28 mm/fraction for left-right direction, from 0.0 to 0.65 mm/fraction for anterior-posterior direction, and from 0.0 to 0.46 mm/fraction for head-feet direction (see Fig. 3.33). Mean slope value was negative for all directions. Values of mean slope were: -0.78 mm/fraction for left-right, -0.57 mm/fraction for anterior-posterior, -0.71 mm/fraction for head-feet direction with standard deviation of 0.78, 0.59 and 0.69 respectively. The greatest observed negative change in time was about -0.35 mm/fraction. In that situation setup error is a decreasing function of time. The biggest observed positive change in time was about 0.65 mm/fraction. In that situation setup error is an increasing function of time. Presented slope values lead to more than 2 cm change in time during 39 fractions of treatment course. In order to check if the absolute setup error value decreases or increases during the course of treatment the difference between the absolute trendline setup error in last and first fraction was examined. Fig. 3.35-3.37 show that for all directions in more than 50% (up to almost 63%) of patients setup errors increased during the course of treatment. Histograms of slope values are presented in Fig. 3.34.

Wilcoxon signed-rank test was used to compare Shapiro-Wilk W statistic values obtained for pure setup errors and obtained for time trend residuals. It showed that, with confidence level set to 95%, we cannot neglect the null hypothesis in left-right and head-feet direction (see Fig. 3.38-3.40). We can interpret this result that at least for this two directions time trends do not explain non-normality of setup errors for some patients. Histograms of Shapiro-Wilk p-values achieved for each patient and direction are shown in Fig. 3.41-3.43.

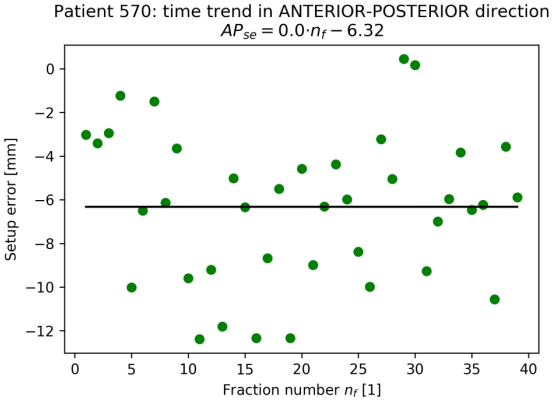
Fig. 3.44 shows that there is no significant difference between number of patients with normal distribution in setup errors in all directions for pure setup errors and time trend residuals.



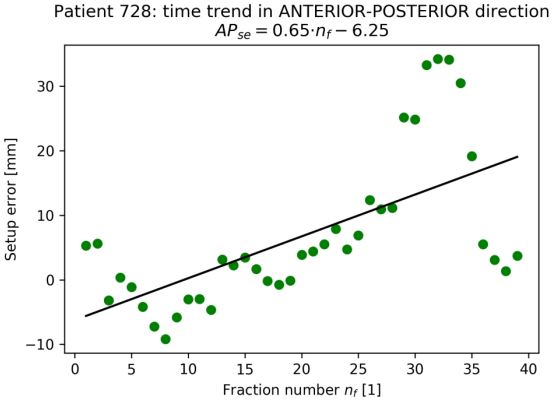
(a) Patient with the lowest observed slope



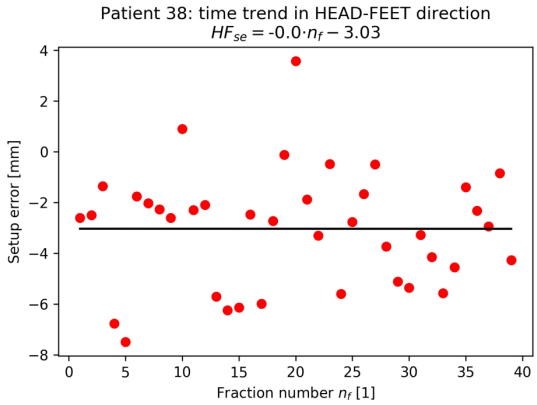
(b) Patient with the highest observed slope



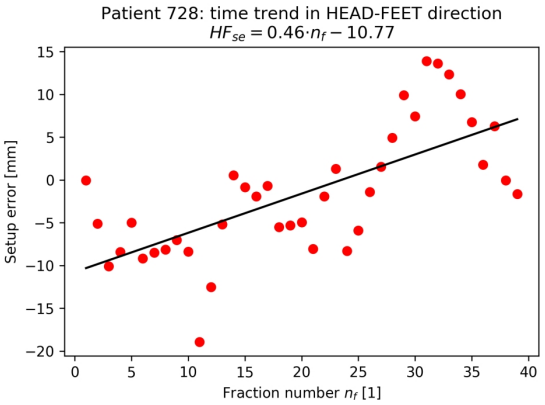
(c) Patient with the lowest observed slope



(d) Patient with the highest observed slope

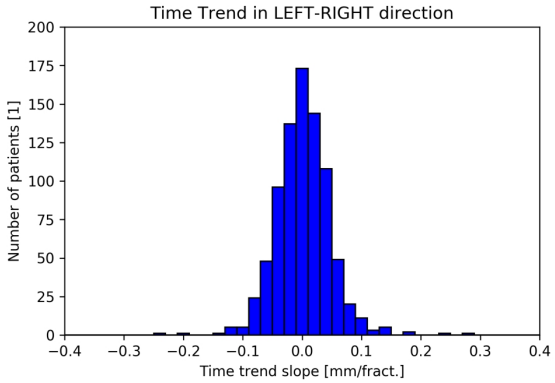


(e) Patient with the lowest observed slope

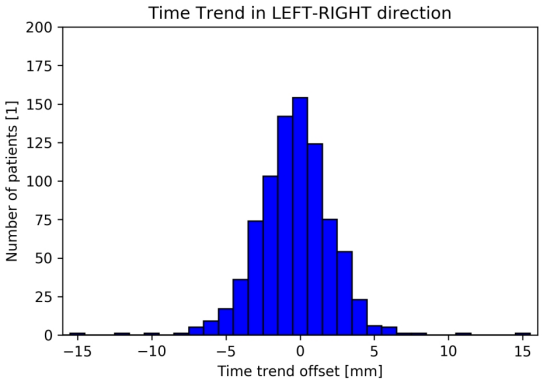


(f) Patient with the highest observed slope

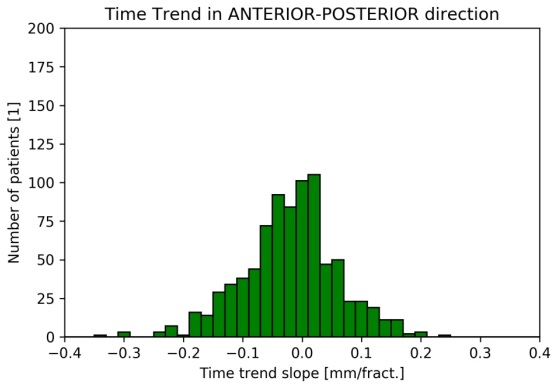
Figure 3.33: Plots of setup error distribution along left-right axis (3.33a and 3.33b), anterior-posterior axis (3.33c and 3.33d) and head-feet axis (3.33e and 3.33f). On the left patients with the lowest observed absolute slope values are shown (with no observed time trend), on the right patients with highest observed slope values are shown (significant time trend observed). LR_{se}, AP_{se}, HF_{se} stands for m_{pf} in particular direction while n_f describes fraction number and is a measure of time. Data achieved for Erasmus MC database.



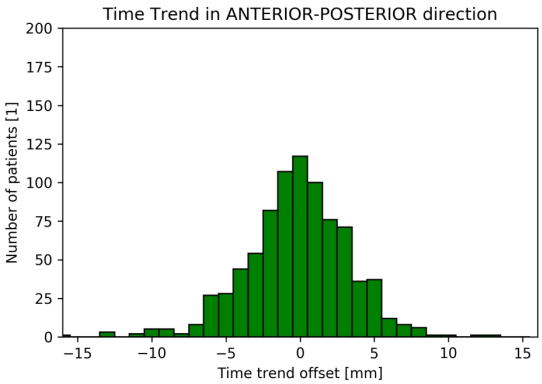
(a) Histogram of slope values



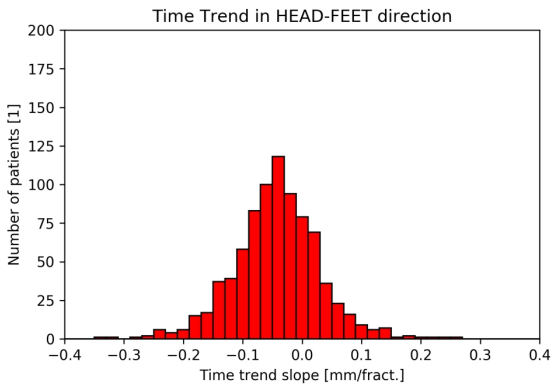
(b) Histogram of offset values



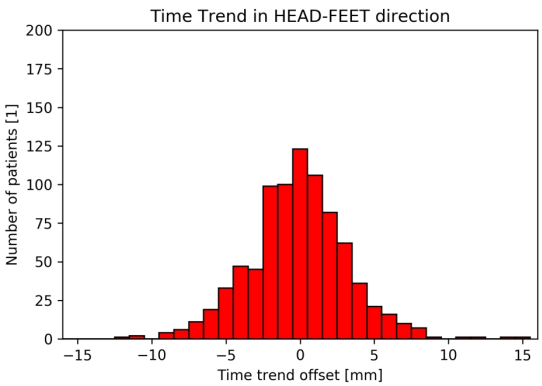
(c) Histogram of slope values



(d) Histogram of offset values



(e) Histogram of slope values



(f) Histogram of offset values

Figure 3.34: Histogram of slope (left) and offset (right) values for time trend OLS analysis of setup errors in each direction. Data achieved for Erasmus MC database.

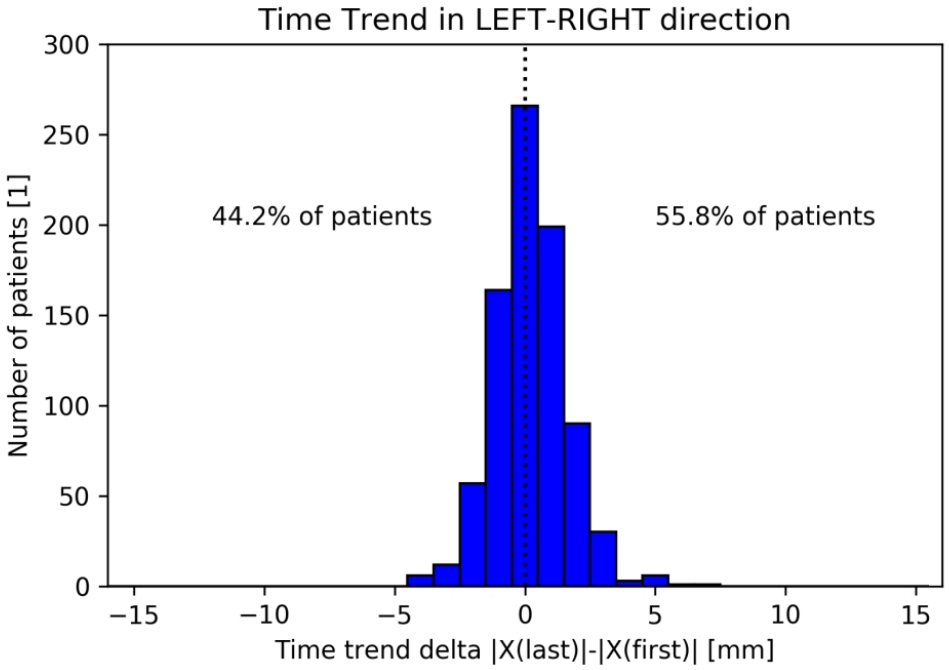


Figure 3.35: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database.

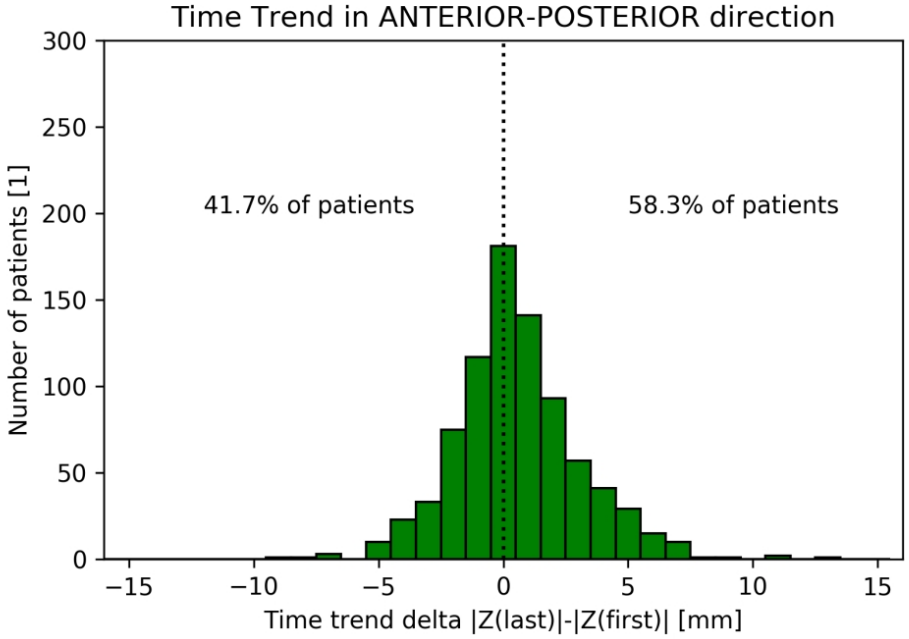


Figure 3.36: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database.

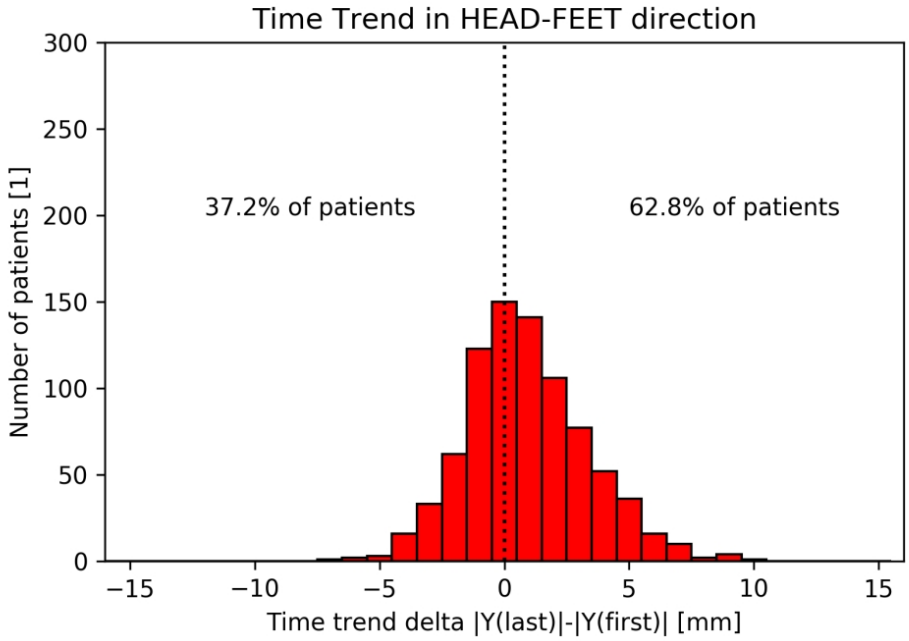


Figure 3.37: Histogram of difference between absolute trendline setup error achieved in last and first fraction. Negative value indicates that setup error got smaller during the course of treatment. Data achieved for Erasmus MC database.

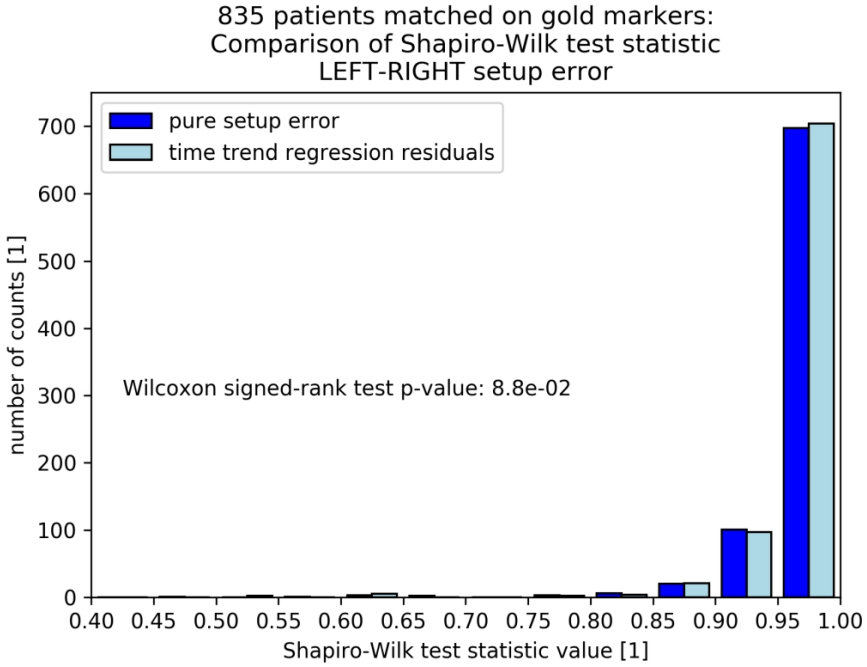


Figure 3.38: Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in left-right direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database.

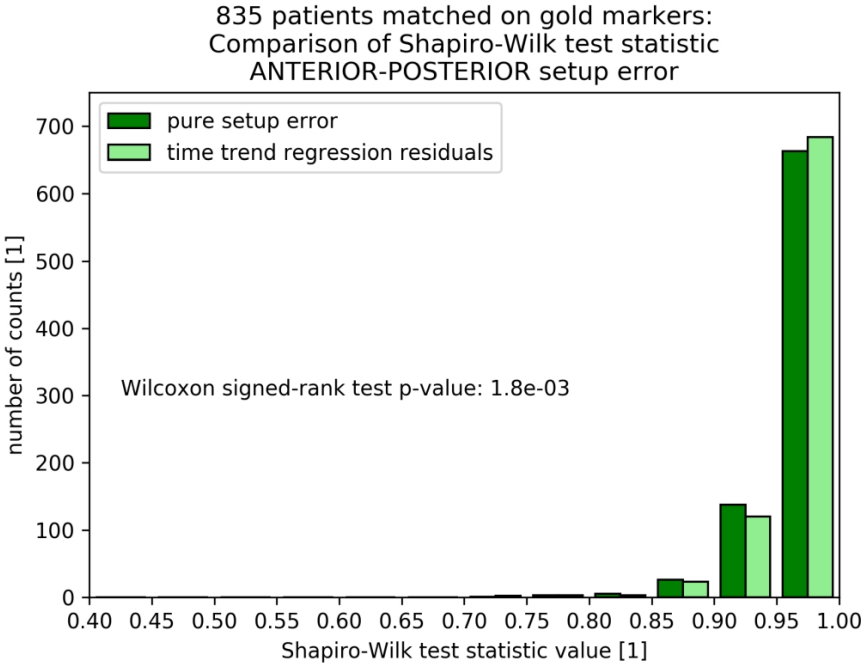


Figure 3.39: Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in anterior-posterior direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database.

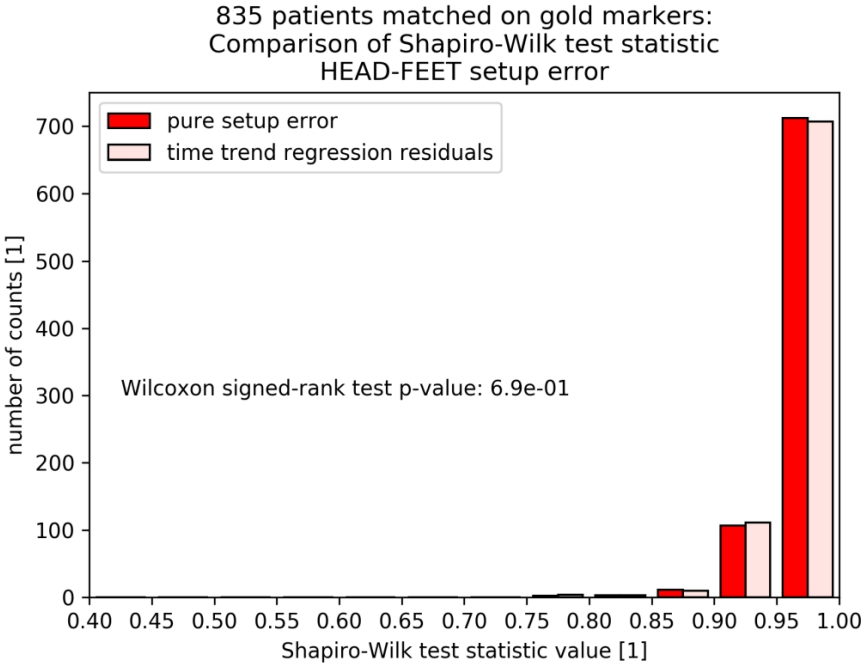


Figure 3.40: Histogram of Shapiro-Wilk statistic W values achieved during normality test of setup errors in head-feet direction for each patient. Histograms for pure setup errors and time regression residuals are shown. Data achieved for Erasmus MC database.

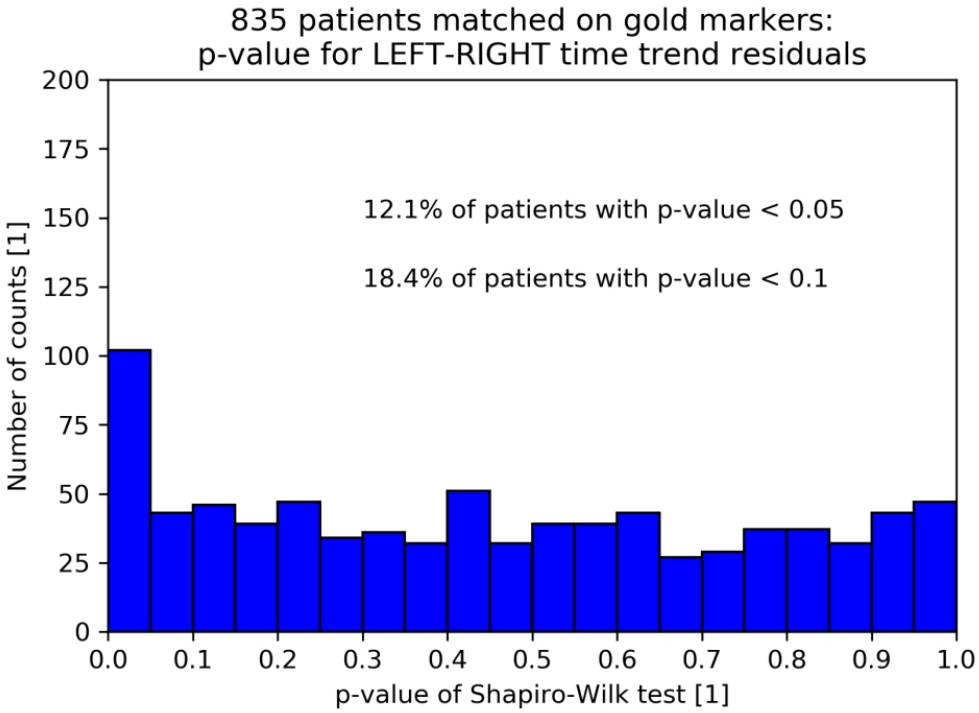


Figure 3.41: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in left-right direction. Data achieved for Erasmus MC database.

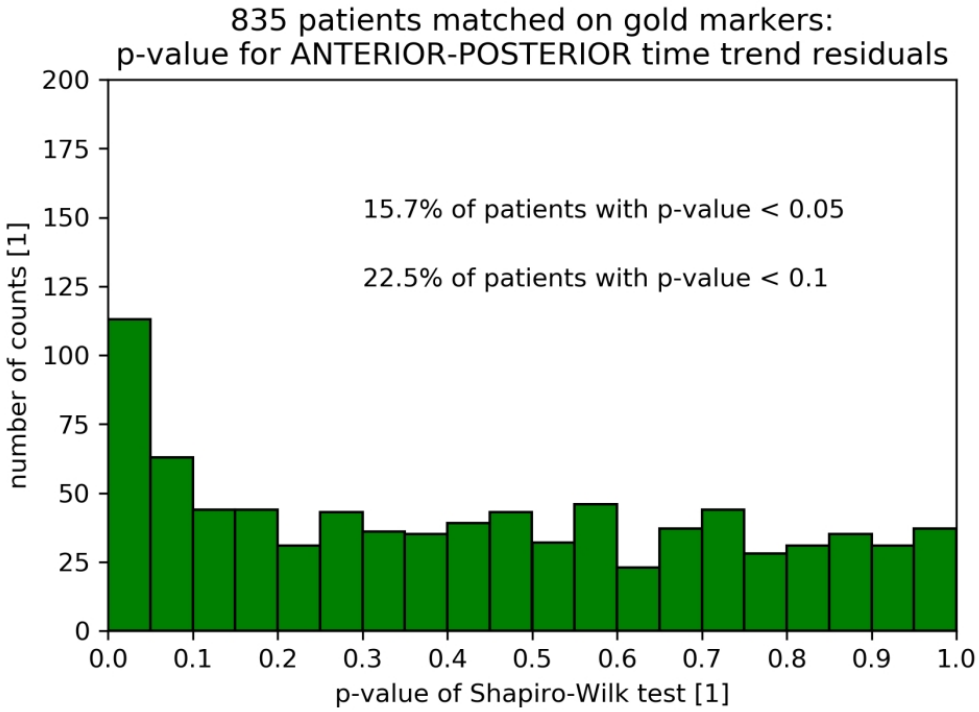


Figure 3.42: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in anterior-posterior direction. Data achieved for Erasmus MC database.

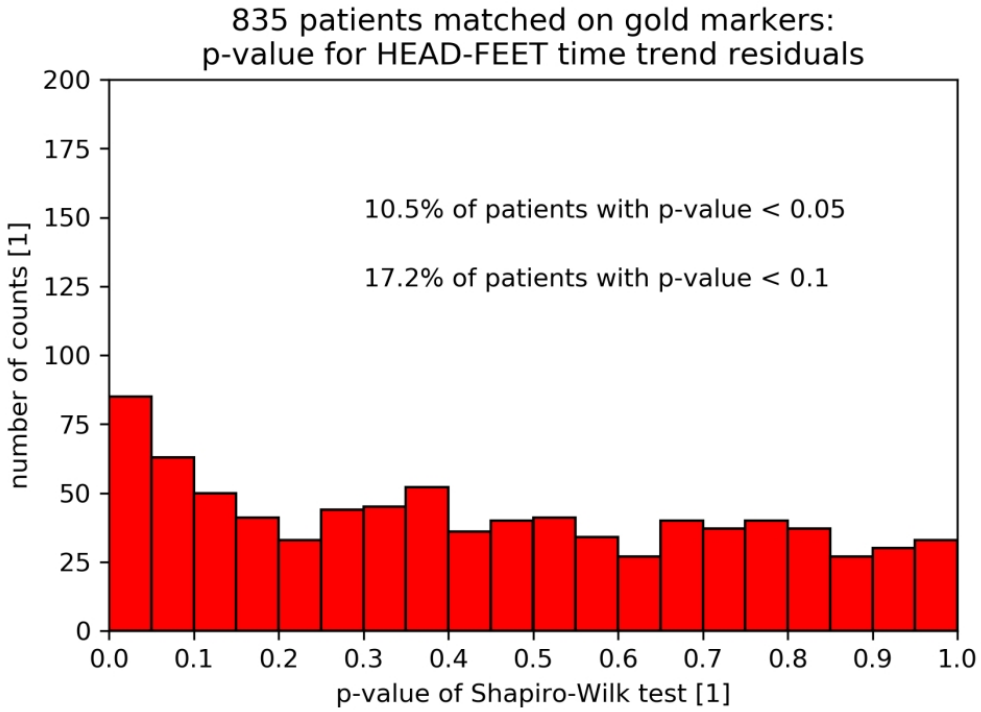


Figure 3.43: Histogram of p-values achieved with Shapiro-Wilk test for time trend residuals in head-feet direction. Data achieved for Erasmus MC database.

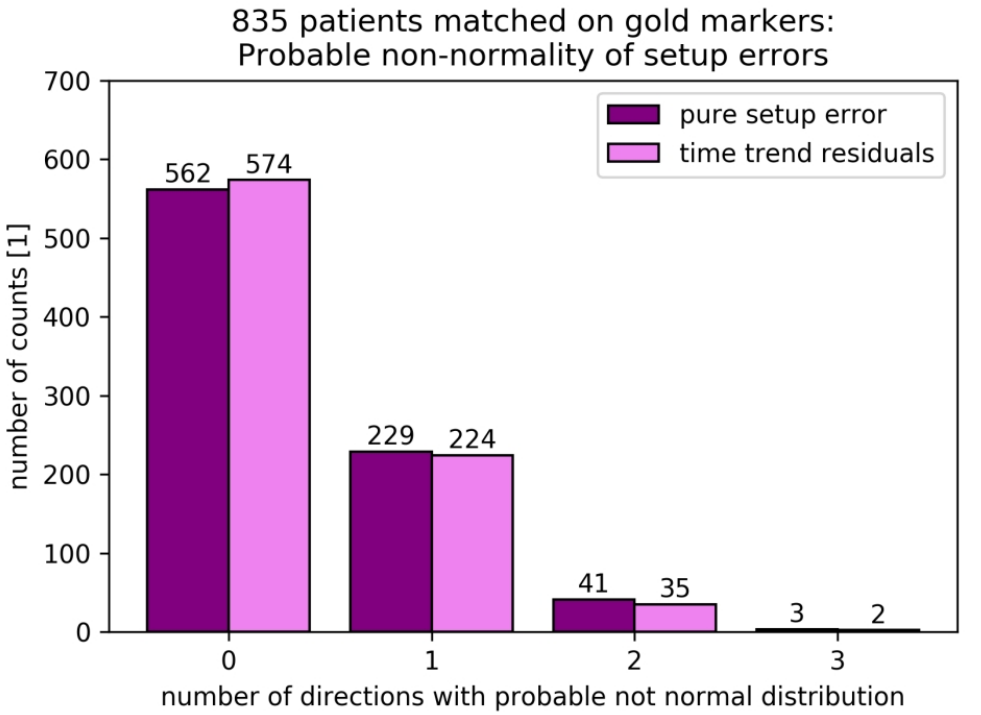


Figure 3.44: Comparison of number of patients with non-normal distribution of pure/time trend residual setup errors in 0-1-2-3 directions. Confidence level was set to 95%. Data achieved for Erasmus MC database.

3.5.3 Fitting probability distributions

3.5.3.1 Mean Setup Error and SD of setup error

Fig. 3.45 presents the distributions of mean setup error and SD of setup errors calculated in three principal directions for Erasmus MC database. The best fitted distribution (in terms of RSS) for mean setup errors was Generalized Logistic distribution with probability density function given by:

$$f(x;\alpha) = \frac{\alpha e^{-x}}{(1+e^{-x})^{\alpha+1}}, \qquad \text{where } \alpha > 0, x > 0 \tag{3.1}$$

The best fitted distribution for standard deviation of setup errors was Burr distribution with probability density function given by:

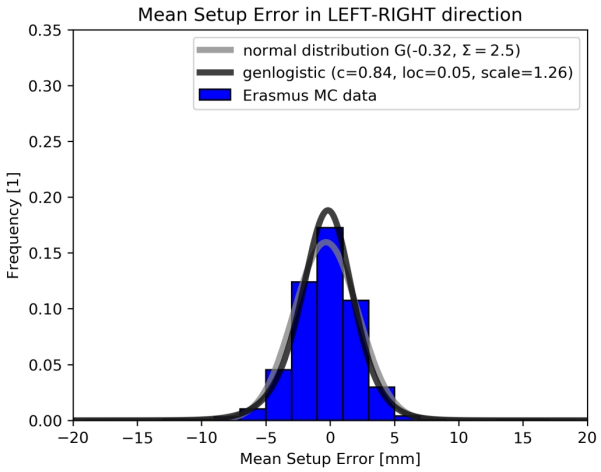
$$f(x;c,d) = c \cdot d \cdot \frac{x^{-c-1}}{(1+x^{-c})^{d+1}} \tag{3.2}$$

Fig. 3.45 presents also normal distribution commonly used in simulation of patient population setup errors. The probability density function for a normal distribution is given by:

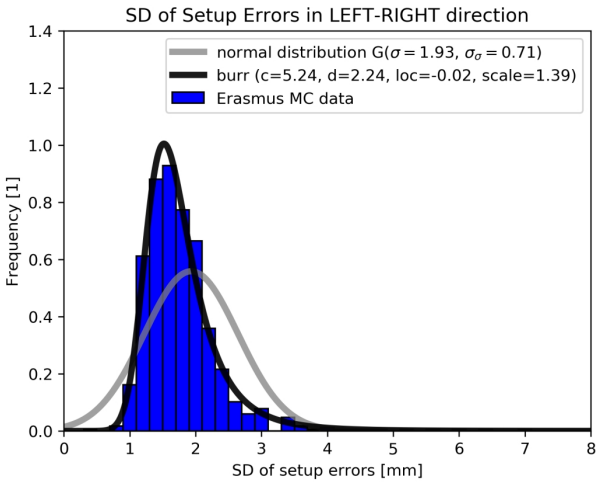
$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \tag{3.3}$$

3.5.3.2 Time Trend Parameters

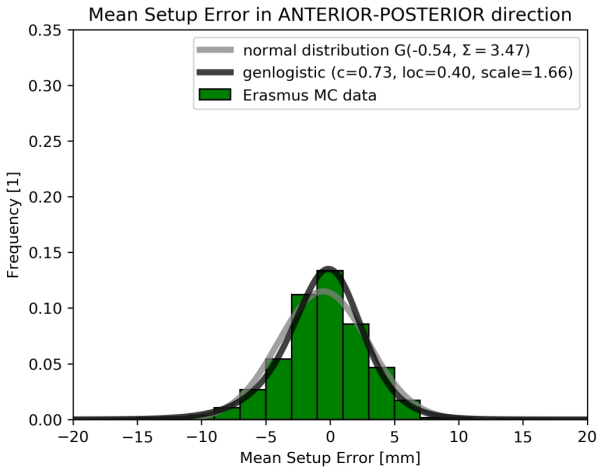
The best fitted distribution for slope and offset of time trends was Burr (Eq. (3.2)) and Generalized Logistic distribution (Eq. (3.1)) respectively – see Fig. 3.46. Standard deviation of residual setup errors SD_p^{tt} have Burr distribution (Eq. (3.2)). Histograms with fitted distributions are presented in Fig. 3.47.



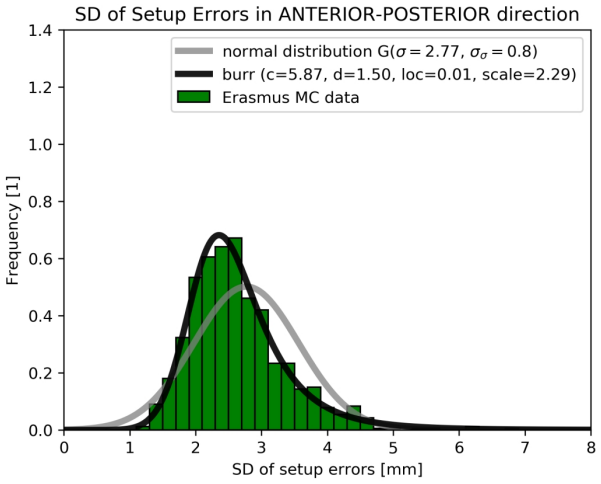
(a) Distribution of $\bar{m}_p$



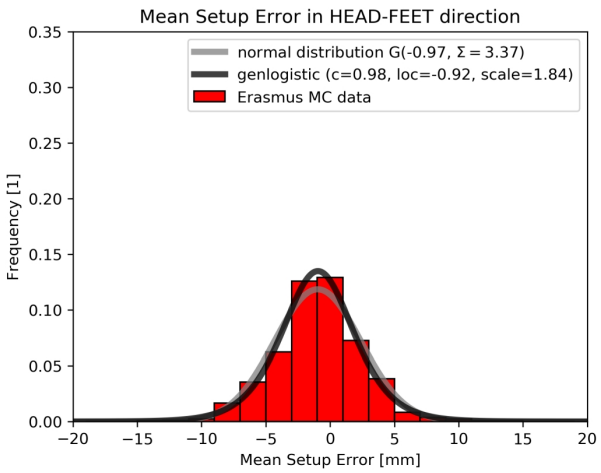
(b) Distribution of SD_p



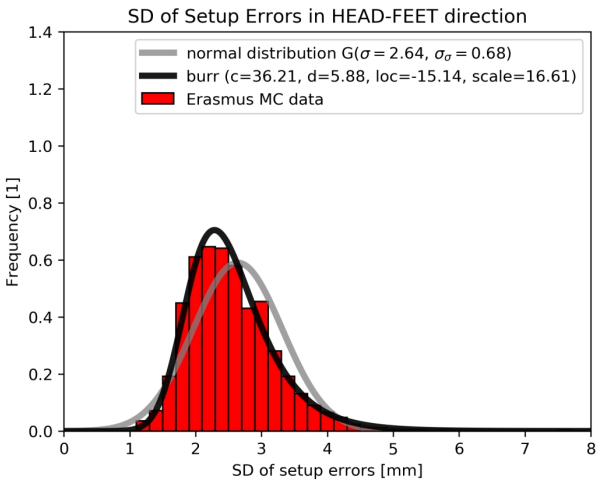
(c) Distribution of $\bar{m}_p$



(d) Distribution of SD_p

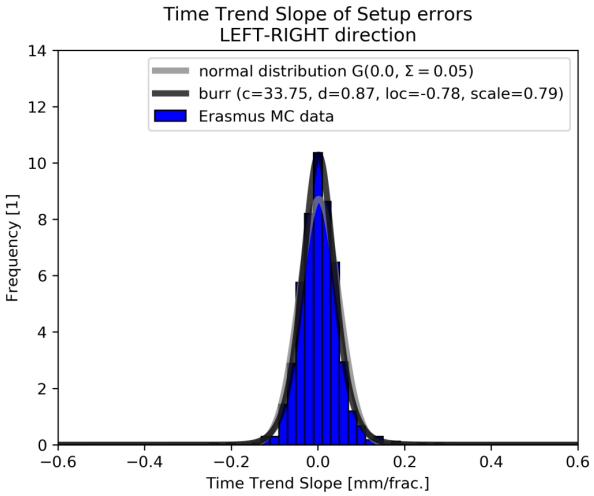


(e) Distribution of $\bar{m}_p$

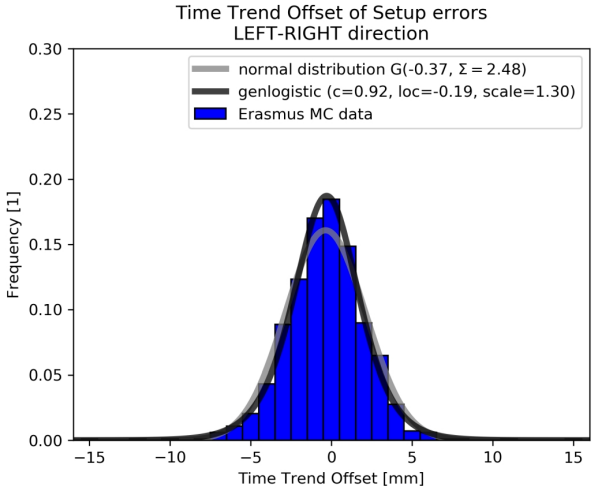


(f) Distribution of SD_p

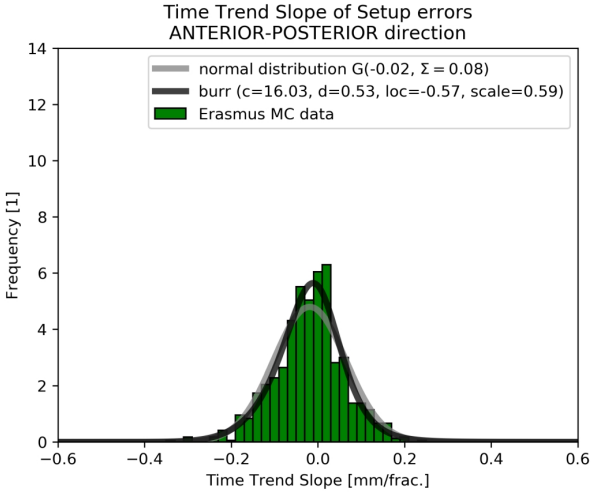
Figure 3.45: Comparison between histograms of mean setup errors and SD of setup errors and fitted distributions. Data achieved for Erasmus MC database.



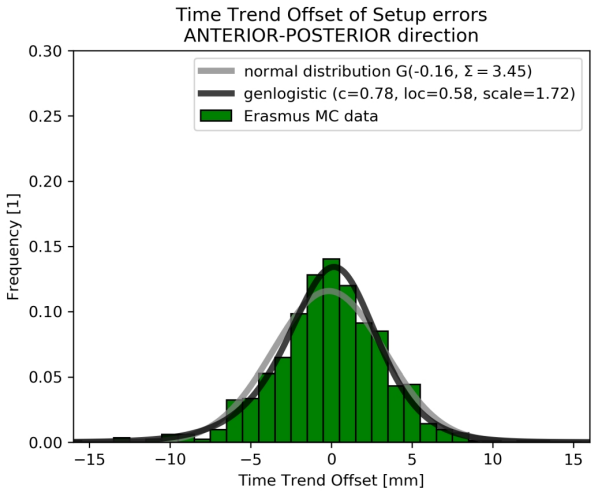
(a) Distribution of a_p



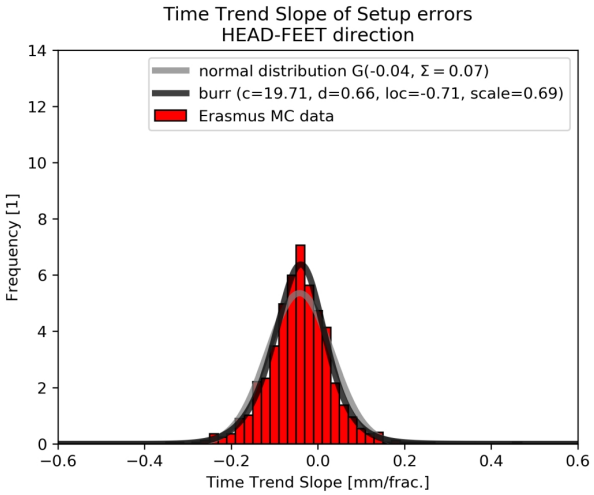
(b) Distribution of b_p



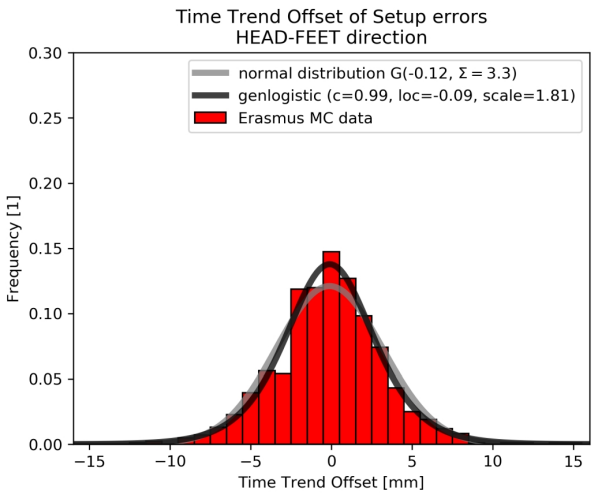
(c) Distribution of a_p



(d) Distribution of b_p

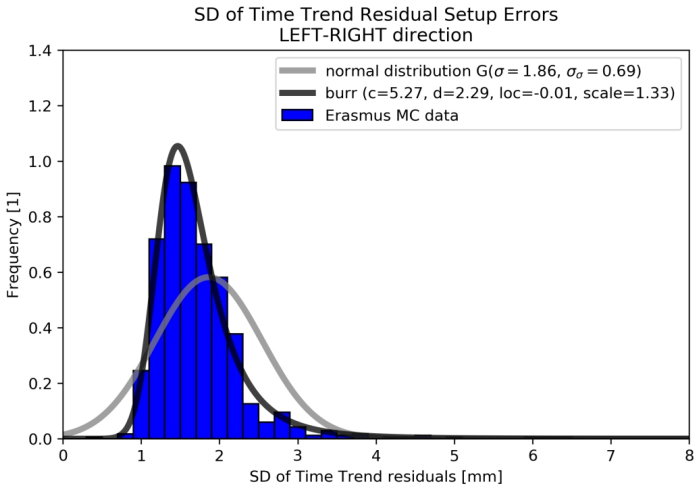


(e) Distribution of a_p

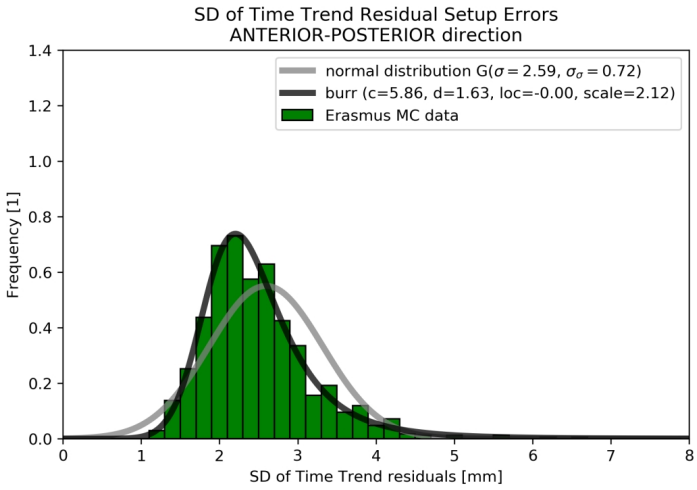


(f) Distribution of b_p

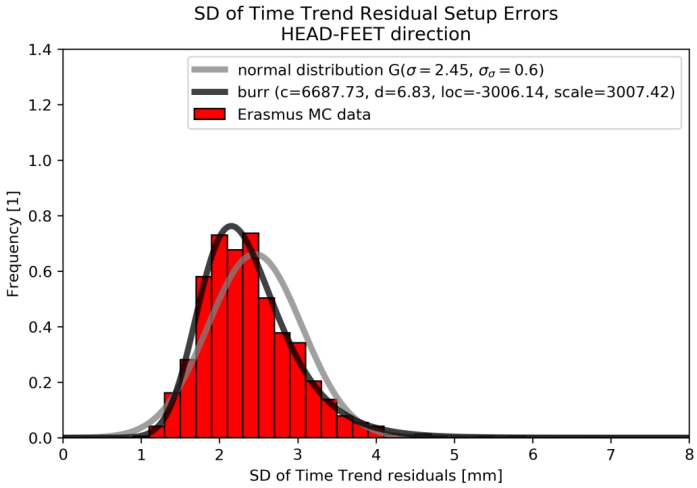
Figure 3.46: Comparison between trendline slope/offset histograms and fitted distributions. Data achieved for Erasmus MC database.



(a)



(b)



(c)

Figure 3.47: Comparison between histograms of SD of Time Trend residual setup errors and fitted distributions. Data achieved for Erasmus MC database.

3.6 Discussion

Analysis of patient data showed that the null hypothesis of normal distribution of setup errors can be rejected for about 40% of patients in case of COI database and for about 20% in case of Erasmus MC database. The difference between both databases is rather big. The possible reasons might be the difference in number of patients in each group (100 vs 835), number of treatment fractions (25 vs 39) and different way of quantifying setup errors (bony anatomy vs gold fiducials implanted in the prostate). Nevertheless for both databases the number of patients with not normal distribution of setup errors is bigger than expected 5% while we use Shapiro-Wilk test with confidence level set to 95%. That results are in contradiction to those presented by Bijhold [4]. It has to be emphasized that although Bijhold concluded the normality of analyzed data it is not clear how tests were done. The only thing presented in his paper, on that matter, are confidence ellipses which seem to present data from normal distribution. In more recent paper Lin et al. [51] investigated normality of prostate motion based on real-time intrafraction tracking. They also performed statistical analysis, with Lilliefors test, for accumulative fractions for 24 patients with at least 30 fractions with measurements. They showed that only 33% of patients had Gaussian distribution in left-right direction, 12.5% in anterior-posterior direction and 8.3% in head-feet direction. My results are based on Shapiro-Wilk test and were performed only for one setup measurement per fraction.

According to Central Limit Theorem one would expect setup errors to have normal distribution. That is because there are many factors influencing patient setup error. All of that errors seem to be of the same order of magnitude and rather small. It is possible that such unpredictable factors like stress (for example related to family and work life) and change in RTTs operating the machine can influence the distribution of setup errors significantly. It has to be emphasized that it is only a hypothesis and therefore should be checked in other works and projects. The importance of the distribution of setup errors is related to the fact that most of protocols and margin recipes used nowadays assume the normal distribution of setup errors [89].

For both databases distribution of SD_p was shown to be not normal. Also distribution of $\overline{m_p}$ in all principal directions for Erasmus MC database and for two directions in COI database was shown to be not normal. This finding and its influence on the methods how we parametrize population of patients was further investigated in Chapter 4.

During radiotherapy we deal with limited number of fractions. That is why it may happen that observed trend has its origin in statistic not in physiology. In order to investigate this issue MC simulation of 10^4 patients, with 40 setup errors each, was created. Four different SD of random errors were used: 1, 2, 3 or 4 mm. Mean systematic error for each patient was taken from the $G(0, \Sigma = 3 \text{ mm})$. Time trend slope was calculated with OLS method using 3 to 40 fractions (see Fig. 3.48). It can be seen that absolute slope value bigger or equal to 0.04 mm/frac can be achieved for 39 fractions treatment with 35% probability. It means that 1.5 mm trend over whole 39 fraction treatment can have a statistic nature.

In order to propose a method of distinguishing between the statistical (apparent) time trend and a physiological one (related to patient body and organ movements) I compared distribution of trend slopes a achieved in synthetic population created without time trend with distribution achieved clinically [34]. Mean slope M_a and width of its distribution Σ_a were compared. The synthetic population consisted of 10^6 patients with $F = 39$ fraction treatment. Setup errors were randomized from Gaussian distributions. Each patient systematic error $\overline{m_p}$ was taken from

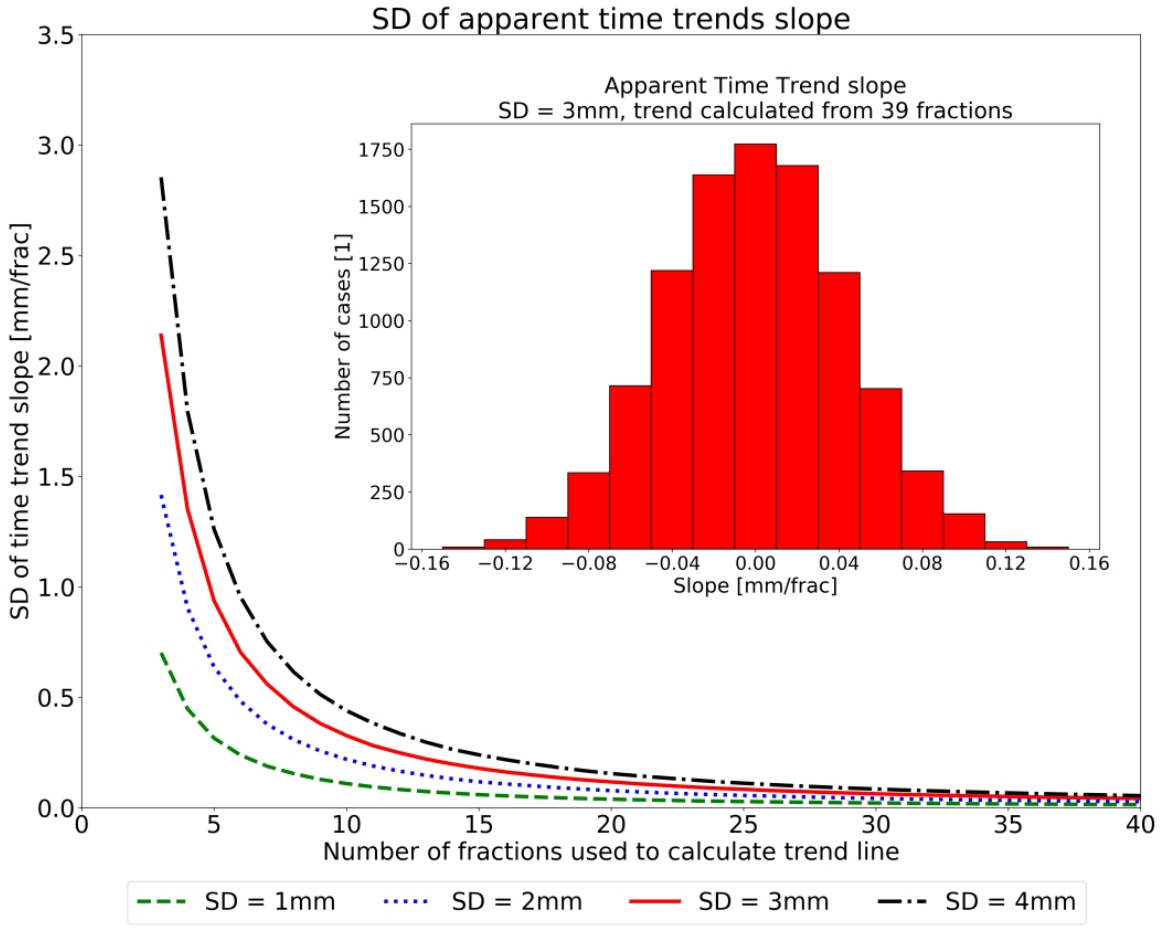


Figure 3.48: SD of slope changing with number of fractions used in OLS trend fit. Histogram of achieved slope values for 39 fractions and $SD = 3$ mm is also shown.

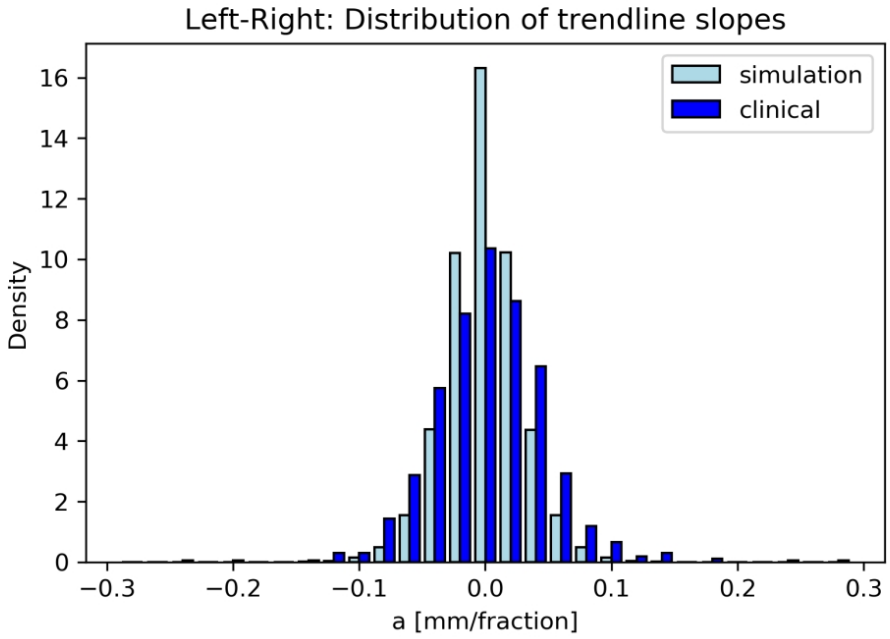


Figure 3.49: Comparison of slope distribution achieved for Erasmus MC database in left-right direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends.

$G(M, \Sigma)$ and SD_p was taken from $G(\sigma, SD_{SD})$. Random errors were taken from $G(\overline{m_p}, SD_p)$. All parameters used in this MC simulation were taken from clinical database analysis (per principal direction). Statistical comparison of synthetic and clinical distribution was done with Kolmogorov-Smirnov test. Fig. 3.49-3.51 show achieved distributions, p-values were $<< 0.001$. Therefore it can be concluded that time trends observed for Erasmus MC database have a physiological component.

Time trend slope of order 0.1 mm/fraction leads to almost 4 mm change during 39 fractions treatment. The observed slope, in both databases, varied up to 0.5-0.6 mm/fraction. The distribution of total trend motion, i.e. difference between first and last fraction according to trendline, for Erasmus MC database is shown in Fig. 3.52.

While analyzing trends, it was also shown, especially for Erasmus MC database, that for more than half of population the setup errors increased during time (see Fig. 3.35-3.37). Usually one would expect that during treatment patient would be more relaxed because he would be more familiar with machine and whole radiotherapy process. This should lead to setup errors decreasing with time. On the other hand there can be so many sources of stress (like medical appointment, change in RTTs staff, machine breakdown, work or family situation, long waiting time for specified fraction) that muscle tension can be hardly predicted. Slope values achieved with OLS method showed greatest values for anterior-posterior direction see Fig. 3.34. That can be explained by muscle tension (see Fig. 3.53). The influence of time trend existence on population parametrization and margin calculation was investigated and would be presented in Chapter 4 and Chapter 7 respectively.

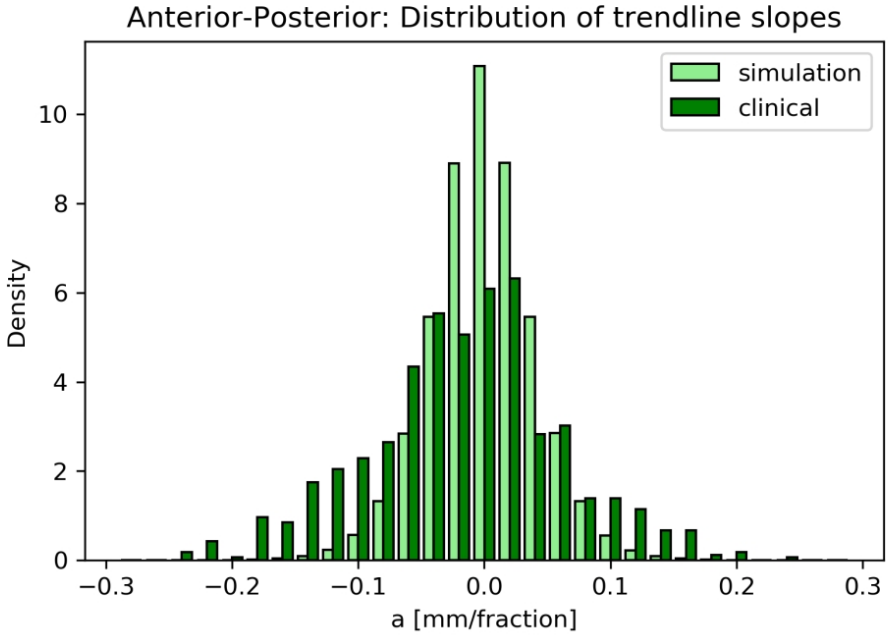


Figure 3.50: Comparison of slope distribution achieved for Erasmus MC database in anterior-posterior direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends.

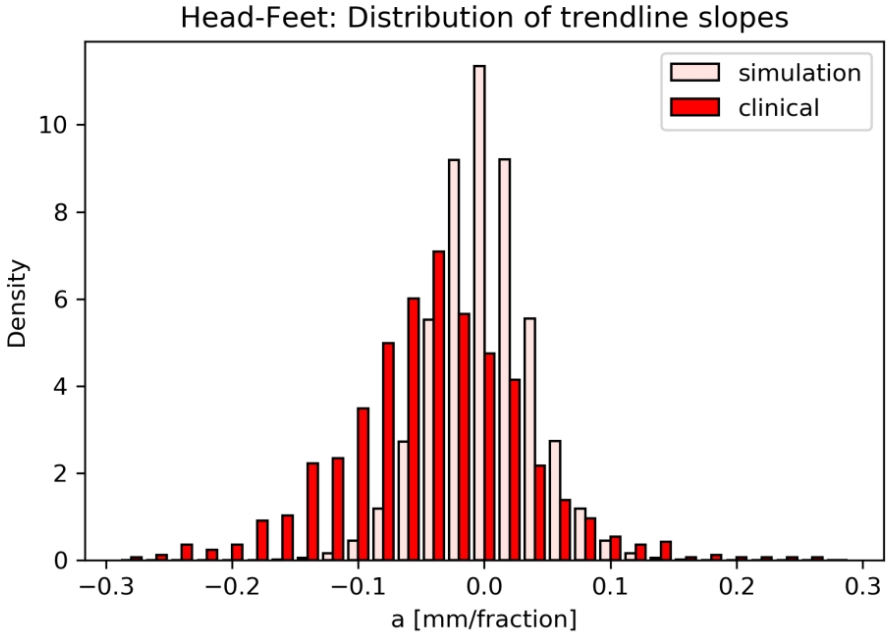


Figure 3.51: Comparison of slope distribution achieved for Erasmus MC database in head-feet direction (clinical) and simulated population of 10^6 patients whose setup errors were not influenced by physiological time trends.

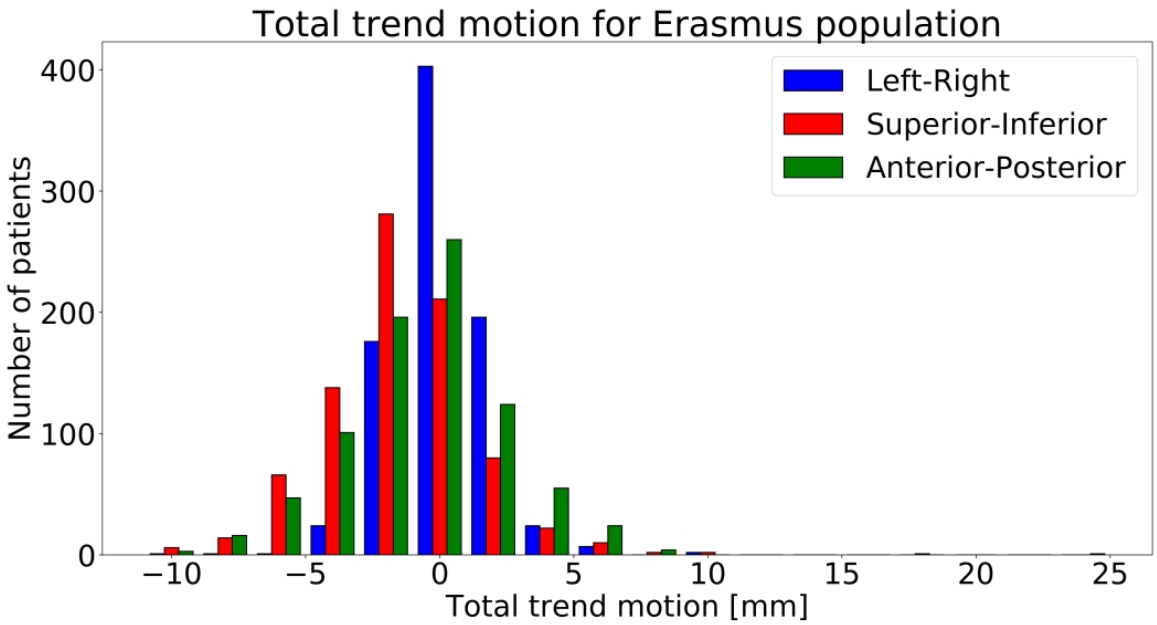
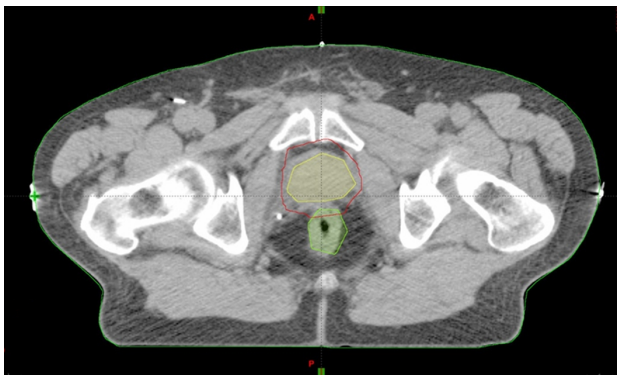
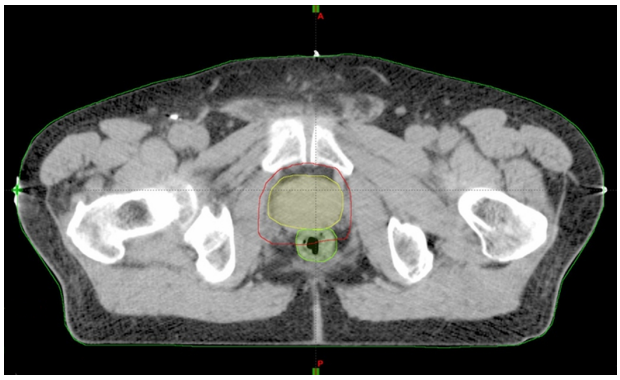


Figure 3.52: Histogram of observed total trend motion (i.e. change in setup errors related only to trendline through the entire treatment). Data achieved for Erasmus MC database.



(a) Tighten buttock muscles



(b) Relaxed buttock muscles

Figure 3.53: Comparison of two CT slices of the same patient. On the left the situation in which muscles' tension can be observed. On the right — relaxed muscles. The change in the PTV position (red line) can be observed — PTV is higher in regard to table top while muscles are tighten. GTV is shown as a yellow region while rectum is a green one.

	σ [**mm**]	M_a [**mm/frac**]	Σ_a [**mm/frac**]
left-right	1.93	0.002	0.046
anterior-posterior	2.77	-0.019	0.083
head-feet	2.64	-0.042	0.075

Table 3.1: Parameters achieved for Erasmus MC database used to check whether observed trends might have a physiological nature.

Chapter 4

Patient population characterization - the choice of right parameters

4.1 Background

The aim of verification protocols is to minimize setup errors and deliver required therapeutic dose to target volume. Probably the best solution is to use on-line verification protocols, i.e. check the patient and target position each day prior to treatment and correct for detected misalignment. The problem of this method is that it is rather time consuming. In clinical practice one has to balance between efficiency and workload. That is why many off-line verification protocols [42] as well as CTV-PTV margins concepts were introduced [89]. Tests of such protocols and margins are mostly done with Monte Carlo (MC) simulations of patient population setup errors data [7, 14, 90]. That is why it is of high importance to have a good patient population parametrization method which would be used during such MC simulations [34]. Otherwise these simulations may lead to erroneous conclusions.

4.2 Materials and methods

In order to check the patient population parametrization methods setup errors of 835 prostate cancer patients treated in Erasmus MC were analyzed. Description of these patient group is given in Section 3.2.2. According to analysis results of the setup data presented in Section 3.5 two methods of population characterization were investigated. One was standard parametrization with Σ to describe systematic errors and σ to describe random errors distribution (see Section 4.2.2.1). The other method incorporated the time trends (see Section 4.2.2.2).

In order to determine the proper way of implementing time trend parametrization pure experiments were done at the beginning. For different simulation parameters setup errors of 10^4 MC patients were simulated. Afterwards these, so called synthetic populations, were analyzed and both parametrizations were applied and compared.

4.2.1 Synthetic population experiments

As it was discussed in Section 3.6 the trend fitted with OLS method to setup errors for treatment with limited number of fractions (even as great as 40) might have only a statistical nature not a physiological one. In consequence for population of patients one can expect to see the coupling

effect of statistical and physiological trends. It has to be determined whether and how to correct for that coupling effect. In order to resolve that problem pure, synthetic populations were created with different parameters using MC simulations of setup errors. Mean population setup error M was set to zero. Σ was taken from $\{1, 2, 3, 4\}$ mm, and σ was taken from $\{1, 2, 3\}$ mm. Width of random error distribution SD_{SD} was set to 0.75 mm. SD of trend slope distribution, Σ_a, was taken from $\{0.0, 0.05, 0.1, 0.15\}$ mm/fraction and mean slope, M_a, was taken from $\{0.0, -0.05, -0.1\}$ mm/fraction. For each set of parameters 39 setup errors (in one direction) were created for 10^4 patients. In total 144 such synthetic populations were generated and used to check and compare parametrization methods in terms of correcting for statistical trends.

On the basis of simulated setup errors classical population parameters were calculated: M, Σ, σ and time trend parameters (M_a, Σ_a – see Section 4.2.2.2).

Assume that we want to reproduce a population with physiological time trend. Time trend in clinical population has both components: statistical and physiological. That is why in order to recreate similar trend in simulation one has to correct for the coupling effect. That can be done by first simulating random errors and calculating slope a_p^r as well as middle position $\overline{m_p^r}$ for them. Afterwards trend defined by a_p^r and $\overline{m_p^r}$ has to be subtracted from random errors and required trend (a_p, $\overline{m_p}$) has to be added. The same applies for standard parametrization in which no trend is considered ($a_p = 0$). As we can always fit a line to randomly chosen errors also mean value, $\overline{m_p^r}$, of them can be always calculated. That mean value, $\overline{m_p^r}$, also has to be subtracted from the random errors. Evaluation of simulated populations was done and is described in Section 4.2.3. On the basis of these pure experiments, methodology of simulating real database was proposed and tested afterwards.

4.2.2 Parametrization methods

Methods used to characterize population of patients in terms of setup errors are described below. In each MC simulation, 10^4 patients with setup errors were generated. For each patient 39 fraction treatment was assumed. Synthetic populations had setup errors only in one direction.

4.2.2.1 Conventional Parametrization

In standard parametrization M, Σ (see Eq. 2.6) and σ with SD_{SD} (see Eq. 2.8-2.7) were obtained from the direct data. In this part of work synthetic populations served as direct data, but clinical data can be used in the same manner. In order to create MC population following steps were performed:

1. Systematic patient setup error, $\overline{m_p}$, was randomized from the normal distribution of mean M and standard deviation equal Σ ($G(M, \Sigma)$).

2. Standard deviation of patient setup errors SD_p was randomized from $G(\sigma, SD_{SD})$ or Burr distribution (see Fig. 3.45)

3. Random setup errors in each fraction were randomized from $G(0, SD_p)$.

4. In order to correct for statistical offset the mean value of random errors was calculated as $\overline{m_p^r}$ and subtracted from random errors.

5. Total setup errors were calculated as sum of systematic error $\overline{m_p}$ and random errors.

4.2.2.2 Trendline Parametrization

In trendline parametrization [16] the time trend slope of patient p — a_p and offset b_p as well as distribution of standard deviations of residual setup errors with $SD_{p,res}$ were obtained for database of real patient data or for synthetic population. The mean setup error for patient p can be calculated as:

$$\overline{m_p} = \frac{1}{F}\sum_{f=1}^{F}(a_p \cdot f + b_p) = \frac{F+1}{2} \cdot (a_p + b_p) \tag{4.1}$$

For total number of fractions $F = 39$ Eq. (4.1) simplifies to:

$$\overline{m_p} = 20 \cdot a_p + b_p \tag{4.2}$$

From Eq. (4.2) it can be concluded that there should be a straight correlation between time trend parameters and mean setup error. Comparison between the distribution of mean setup error calculated directly from setup errors and from time trend parameters is shown in Fig. 4.1-4.3. The correlation between time trend parameters and mean setup error was also investigated (see Fig. 4.4). The R^2 parameter and p-value were calculated. The R^2 parameter gives information of how well data can be explained by the model (here linear fit). The p-value gives the information of F statistic testing the null hypothesis that intercept-only model and linear fit are equal. From results presented in Fig. 4.4 it can be concluded that the null hypothesis of correlation between the mean setup error and time trend offset value cannot be neglected. The linear correlation between the time trend slope and mean setup error is also significant (p-value below 0.05) but not so visually clear like the one for offset.

In order to create MC populations with time trend parametrization following steps were performed, assuming that required parameters were previously calculated for direct data:

1. Middle position of patient setup error, $\overline{m_p}$, was randomized from the normal distribution of zero mean and standard deviation equal Σ ($G(M, \Sigma)$). Note that in order to emphasize the difference between middle position and slope distribution m superscript was added: $M = M_m$ and $\Sigma = \Sigma_m$.

2. Patient slope a_p was randomized from $G(M_a, \Sigma_a)$

3. Standard deviation of patient residual setup errors SD_p^{tt} was randomized from $G(\sigma^{tt}, SD_{SD}^{tt})$ or Burr distribution (see Fig. 3.47)

4. Random setup errors in each fraction were randomized from $G(0, SD_p^{tt})$.

5. In order to correct for statistical trend the OLS method was used to determine the trend of random errors yielding to slope a_p^r and mean error $\overline{m_p^r}$. Afterwards this trend was subtracted from residual random errors.

6. The patient trend was calculated as: $(f - \frac{F+1}{2}) \cdot a_p + m_p$, with f describing each fraction of treatment.

7. Total setup errors were calculated as sum of patient trend and residual random errors.

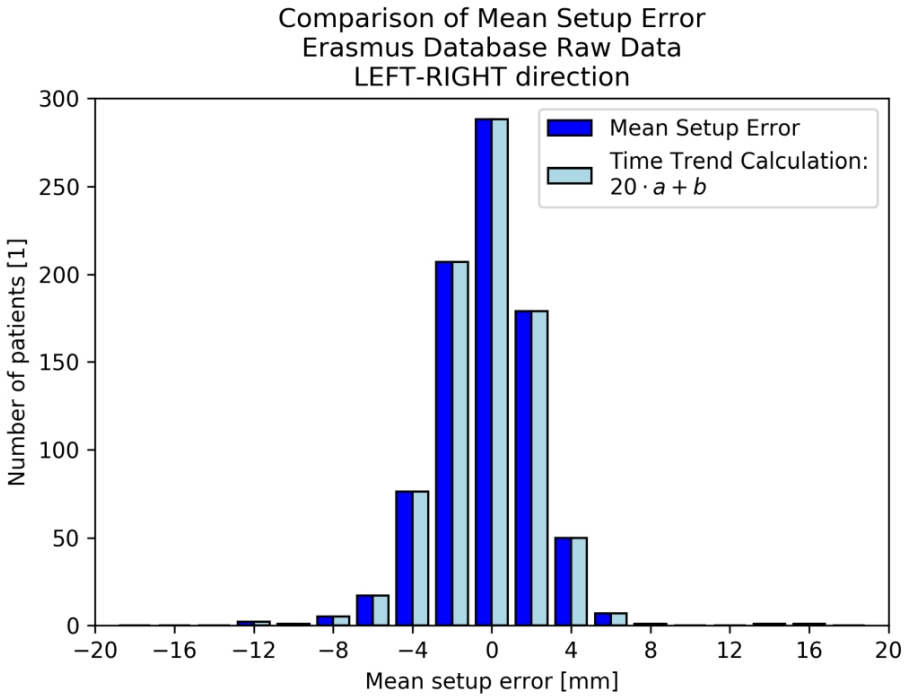


Figure 4.1: Histogram of left-right mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database.

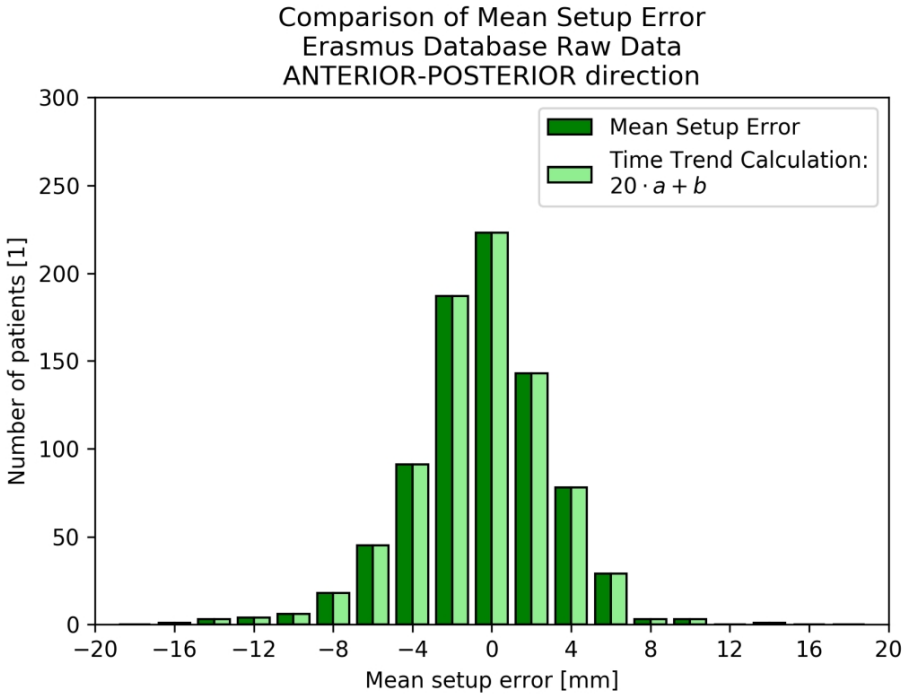


Figure 4.2: Histogram of anterior-posterior mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database.

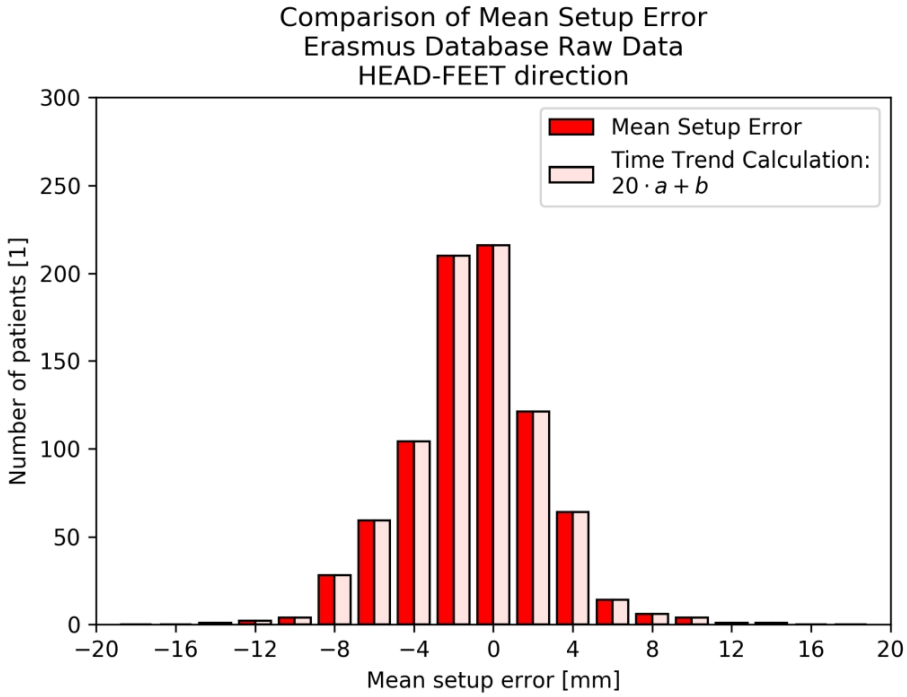


Figure 4.3: Histogram of head-feat mean setup error calculated directly from database as well as calculated from the time trend analysis parameters. Data achieved from the Erasmus MC database.

4.2.3 Evaluating quality of parametrization

4.2.3.1 Ability of reproducing the initial data

In order to check the efficiency of the proposed parametrization methods distributions of parameters used to simulate the population were evaluated. It was assumed that a good parametrization should reproduce the inserted data. The distribution of mean setup error and standard deviation of setup error were compared at the beginning. The distribution of time trend slope and offset were also compared. Additionaly, for Erasmus MC database, the d_{100} defined as a distance within each 100% of absolute setup errors lied in, was also evaluated. Two-sample Kolmogorov-Smirnov test was used to compare distributions of input population with recreated one (MC population). Null hypothesis stated that there is no difference between the two and it was neglected if the probability of achieving statistic value as observed for two similar distributions was lower than 0.001 (p-value ≤ 0.001). The choice of such low p-value was based on many parameters influencing final distribution of setup errors and their parameters.

4.2.3.2 NAL and eNAL verification protocols performance

In order to check and compare population characterization methods NAL verification protocol (Section 2.3.1) and eNAL verification protocol (Section 2.3.2) were applied for the clinical and synthetic populations. NAL protocol was applied for different number of fractions taken into the correction calculation (range 3 to 39). eNAL protocol was applied with starting 3 fractions with imaging and a correction calculated afterwards and updated weekly in the remaining treatment according to the fitted trend. The residual Σ (Σ_{res}, see Eq. 2.6) was calculated

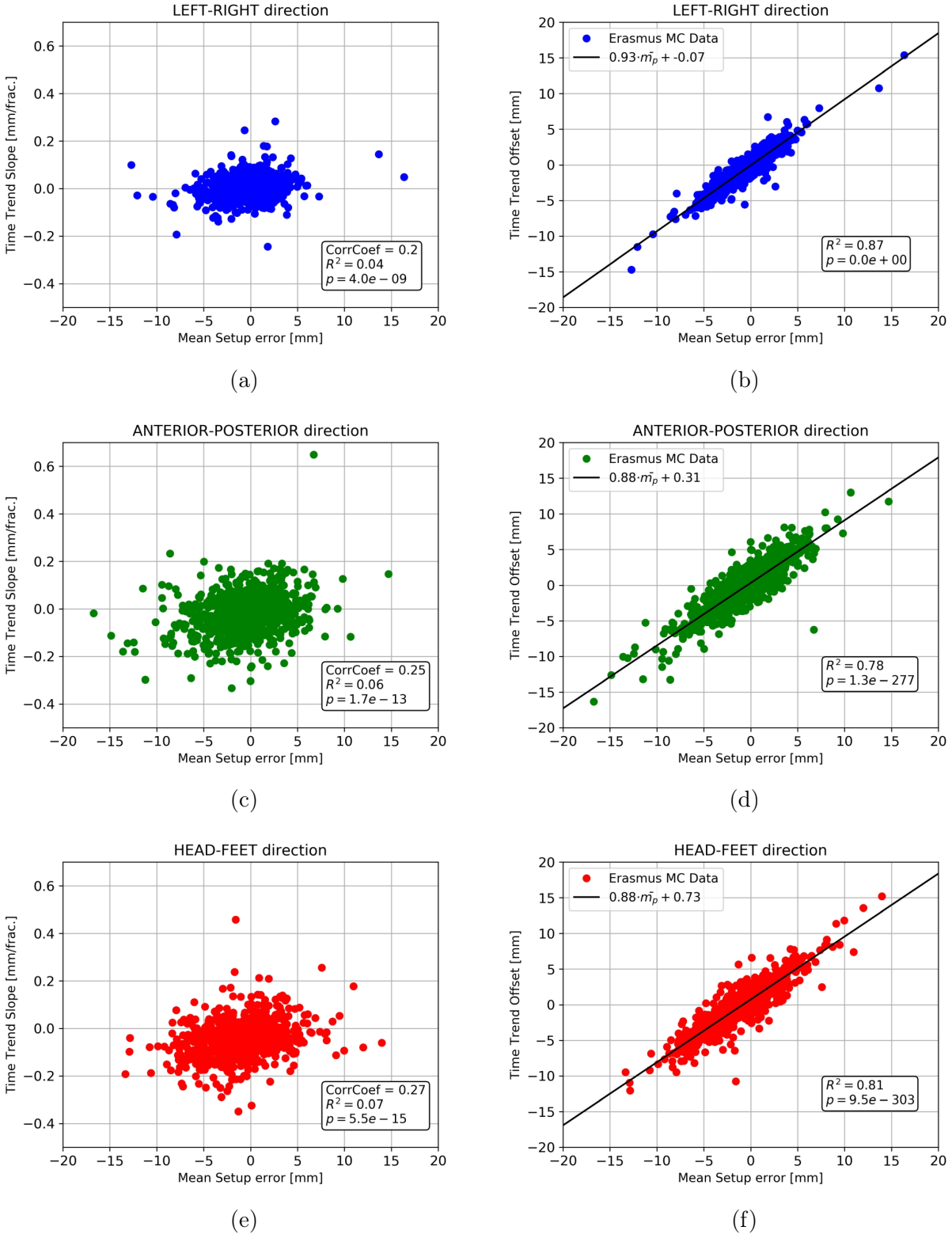


Figure 4.4: Correlation between slope (left) and offset (right) for time trend OLS analysis and mean value of setup errors in each direction. Data achieved for Erasmus MC database.

each time and compared between the methods used to parametrize the patient population and perform MC simulations. The hypothesis was that a good parametrization should allow to predict NAL/eNAL protocol performance of inserted/based data. In case of NAL protocol it means that relation between Σ_{res} and number of fractions used to calculate the correction factor should be the same.

4.3 Results

4.3.1 Synthetic population experiments

4.3.1.1 Ability of reproducing the initial data

For both simulation methods, i.e. using the conventional parametrization (conventional MC) and the trendline parametrization (trendline MC), distributions of $\overline{m_p}$ fully reproduced distributions of synthetic populations (p-value > 0.001). SD_p was significantly different for all conventional MC populations and for most of trendline MC populations. Only 4 trendline populations reproduced SD_p distribution properly. Time trend slope (a_p) distributions were significantly different from the direct population for 122 out of 144 conventional MC populations. For trendline MC populations the slope was fully reproduced in all cases. Time trend offset was significantly different for all conventional MC populations and for 36 trendline MC populations.

4.3.1.2 NAL and eNAL verification protocols performance

Example results of NAL verification performance for synthetic and MC populations are shown in Fig. 4.5. Despite the parameters used for the synthetic population generation, Σ_{res} for direct simulation was well reproduced by trendline MC (in fact the direct simulation and trendline MC curves lies on each other). It was not the case for the conventional MC. For most of the cases with a trend present the conventional MC underestimated Σ_{res}. For populations with no trend the conventional MC overestimated Σ_{res}. For some populations having trendline parameters comparable with those related to limited number of fractions the conventional simulation could properly reproduce Σ_{res}.

Fig. 4.6 presents results of Σ_{res} for $n = 3$ and different parameters of synthetic populations (M_a, Σ_a, σ'). Presented results does not depend on value of Σ used for synthetic population creation. For $\Sigma_a = 0$, despite other parameters used, conventional parametrization overestimates Σ_{res}. It might be explained by the fact that for $\Sigma_a = 0$ and $M_a < 0$ with increasing number of fractions setup error was decreasing leading to smaller value of Σ_{res}. That behavior cannot be properly modeled with conventional parametrization. For $\Sigma_a = 0$ and $M_a = 0$ there are still trends related to limited number of fractions which were not corrected for in conventional parametrization. With increasing Σ_a, Σ_{res} calculated for conventional parametrization gets closer to direct simulation (synthetic population) depending on σ' and M_a. For bigger values of Σ_a conventional parametrization underestimates Σ_{res}. Using conventional parametrization can lead to error in predicting Σ_{res} even of 1-1.5 mm. Trendline parametrization corrected for limited number of fractions is the best to model synthetic population with differences in Σ_{res} not exceeding 0.1 mm.

In Fig. 4.7 eNAL performance is compared between conventional and trendline parametrizations. Similarly as for NAL protocol the trendline parametrization is in agreement with direct

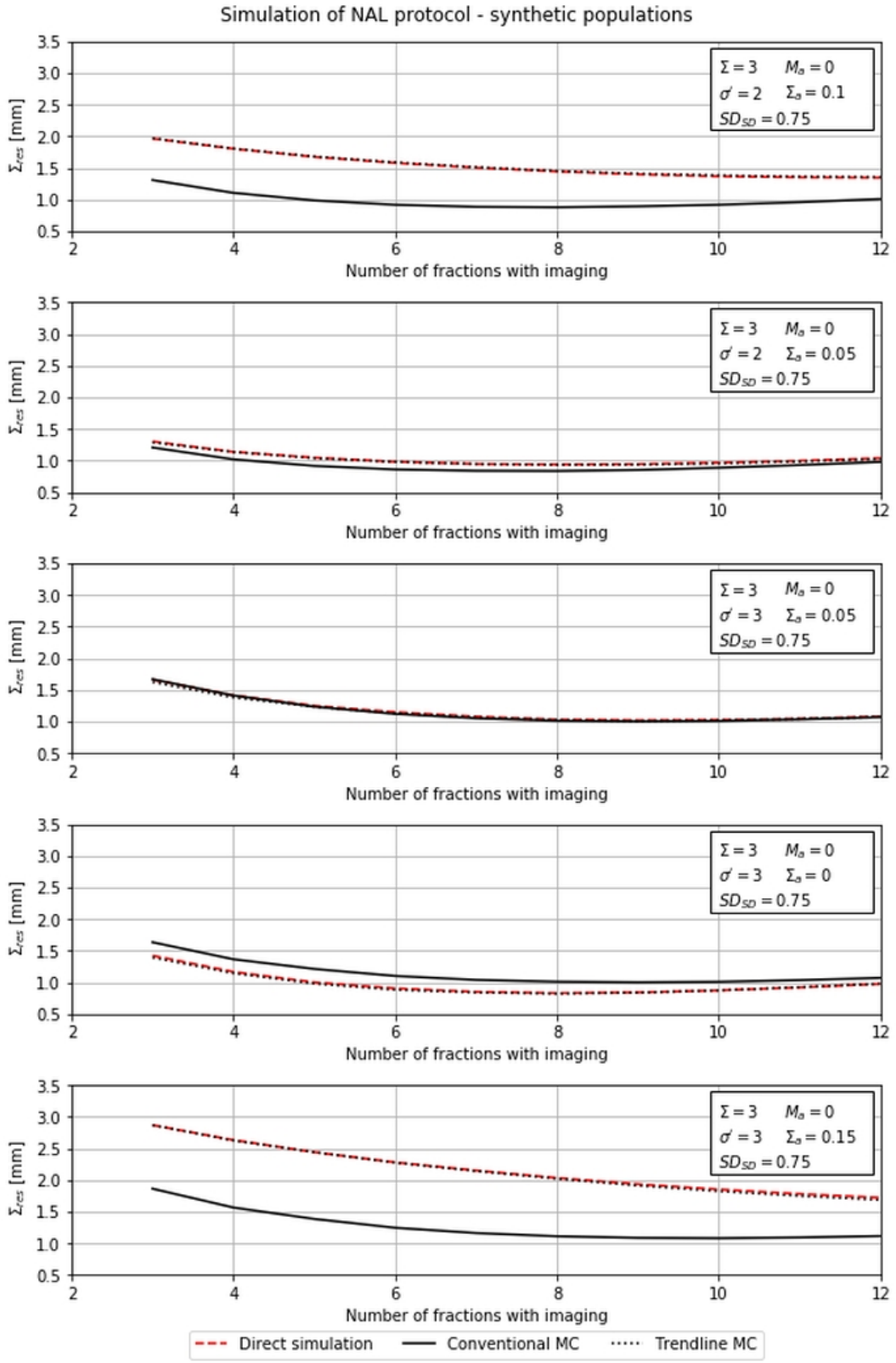


Figure 4.5: Comparison of Σ_{res} achieved with NAL protocol for five arbitrary chosen synthetic and MC populations. Direct simulation shows the NAL performance for synthetic population, it is ground truth for this evaluation.

simulations (ground truth achieved here for synthetic populations). Unlike for NAL protocol the conventional parametrization overestimates Σ_{res} which may lead to application of too big margins.

4.3.2 Clinical Data

4.3.2.1 Ability of reproducing the initial data

Distributions of mean setup error $\overline{m_p}$, SD of patient setup errors SD_p, d_{100}, patient trend slope a_p and patient trend offset b_p were investigated and compared between clinical data (Erasmus MC database) and performed MC simulations. In total four simulation methods were compared at this step. In general two methods of parametrization were used: conventional parametrization (4.2.2.1) and trendline parametrization (4.2.2.2). Each of the parametrization methods was corrected for limited number of fractions. For both methods two distributions of SD_p were tested for random errors generation: Gauss (see Eq. 3.3) and Burr (see Eq. 3.2).

Table 4.1 shows summary of achieved comparisons between performed MC simulations and clinical data in terms of achieved p-values. If these p-values were bigger than 0.001 null hypothesis that there is no difference between compared distributions couldn't been neglected. These values are presented in bold. It can be seen that in order to reproduce distribution of SD_p and d_{100} Burr distribution should be used to randomize SD of patient random errors. To reproduce slope and offset distribution trendline parametrization should be used. Detailed results for both parametrization methods are given in following two paragraphs.

	$\overline{m_p}$	SD_p	d_{100}	a_p	b_p
Conventional Paramentrization (Gauss)					
Left-Right	**0.249**	0.0	0.0	0.0	**0.031**
Anterior-Posterior	**0.021**	0.0	**0.286**	0.0	0.0
Head-Feet	**0.101**	0.0	**0.034**	0.0	0.0
Conventional Paramentrization (Burr)					
Left-Right	**0.338**	**0.116**	**0.738**	0.0	**0.058**
Anterior-Posterior	**0.006**	**0.016**	**0.74**	0.0	0.0
Head-Feet	**0.176**	**0.079**	**0.275**	0.0	0.0
Trendline Paramentrization (Gauss)					
Left-Right	**0.265**	0.0	0.0	**0.174**	**0.022**
Anterior-Posterior	**0.011**	0.0	**0.002**	**0.016**	**0.003**
Head-Feet	**0.174**	0.0	**0.015**	**0.042**	**0.009**
Trendline Paramentrization (Burr)					
Left-Right	**0.441**	**0.899**	**0.469**	**0.105**	**0.032**
Anterior-Posterior	**0.03**	**0.493**	**0.586**	**0.041**	**0.003**
Head-Feet	**0.086**	**0.368**	**0.208**	**0.052**	**0.002**

Table 4.1: Table presents p-values of Kolmogorov-Smirnov for comparison of clinical database distributions with different methods of parametrizing and simulating setup errors. In brackets distribution used in MC simulation for selection of SD_p is given. p-values ≥ 0.001 are bold and indicate for which parameter distribution achieved with tested parametrization was comparable to clinical data (i.e. null hypothesis could not been rejected).

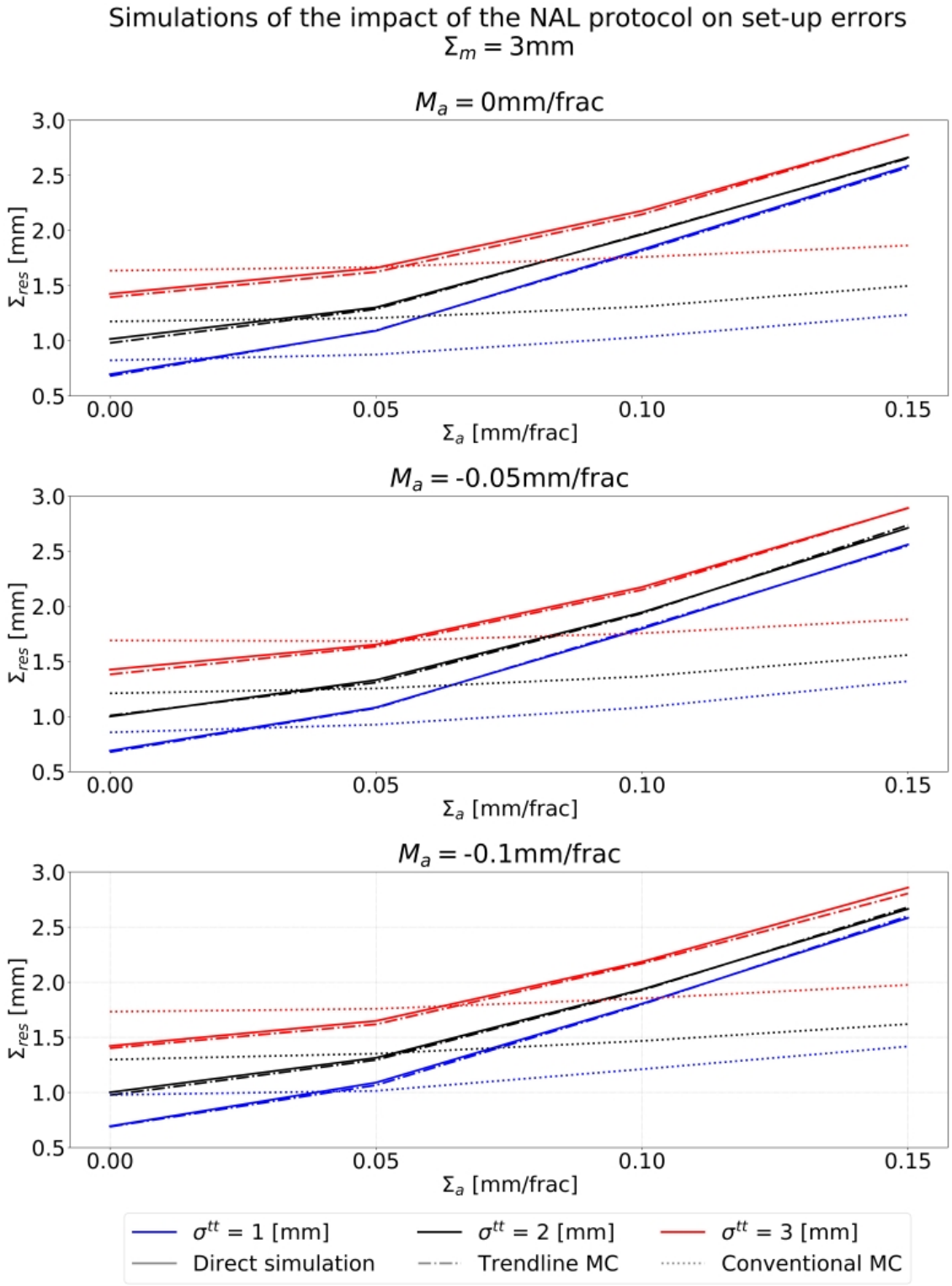


Figure 4.6: Values of Σ_{res} achieved for 36 different synthetic populations (direct simulation) and adequate MC populations: with standard parametrization (Conventional MC) and trendline parametrization (Trendline MC). In each case NAL protocol was applied after first 3 fractions with imaging. Total treatment time was set to 39 fractions.

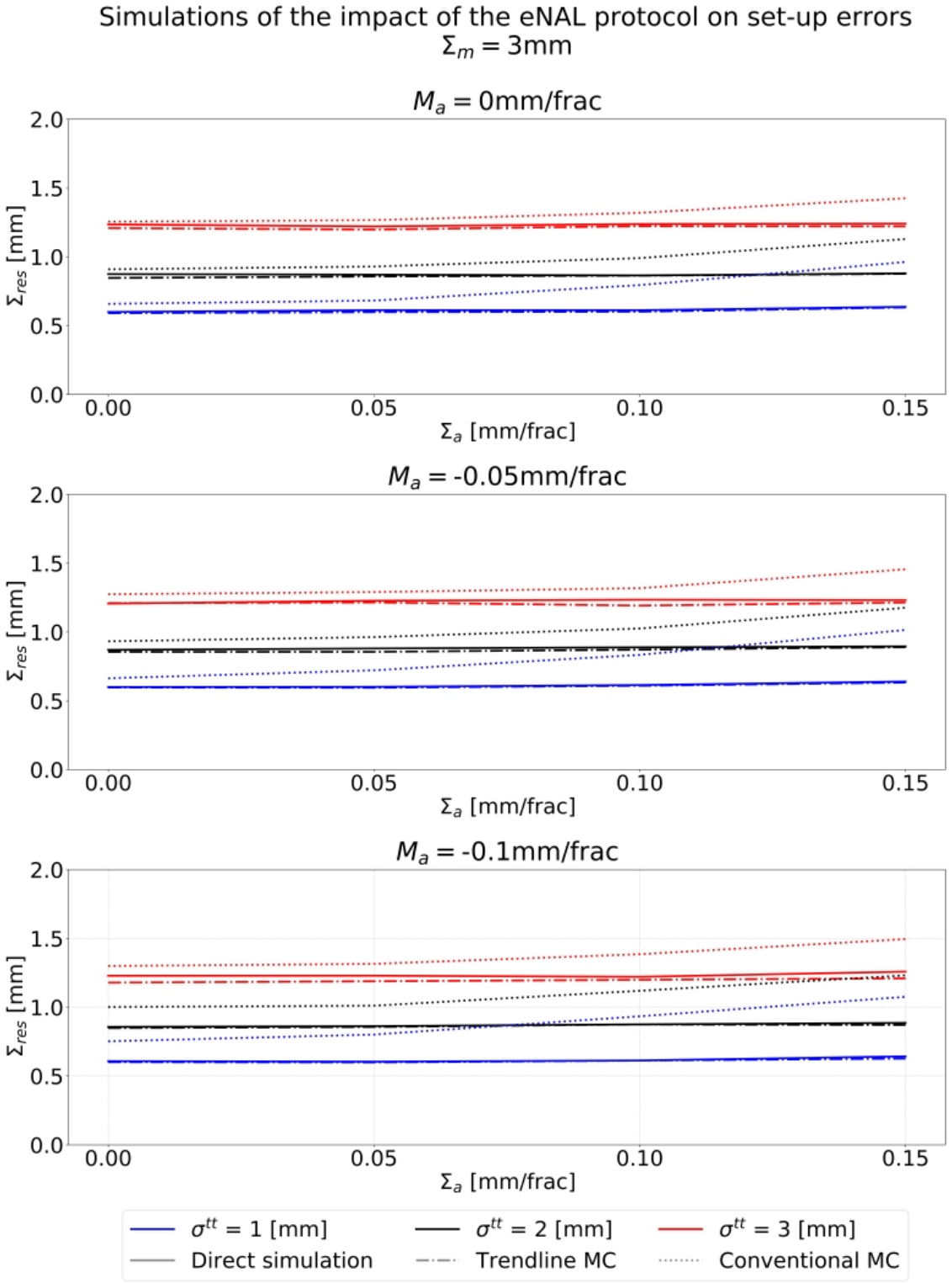


Figure 4.7: Values of Σ_{res} achieved for 36 different synthetic populations (direct simulation) and adequate MC populations: with standard parametrization (Conventional MC) and trendline parametrization (Trendline MC). In each case eNAL protocol was applied in which first correction was done after third fraction with update of correction factor done weekly. Total treatment time was set to 39 fractions.

Conventional Parametrization

Fig. 4.8-4.12 present comparison of distributions for parameters used to evaluate and parametrize the database of setup errors between real data and MC simulations with conventional parametrization. It can be clearly seen that there is a significant difference between the real database and simulation in distribution of SD of random errors while using Gaussian function to randomize SD_p. Also d_{100} is better reproduced with Burr distribution used for random errors generation. With conventional parametrization it is impossible to reproduce slope and offset distribution as it was calculated for real data. Still some distribution of slope would be observed according to limited number of fractions. The slope calculated from data taken randomly from normal distribution with 95% confidence level has an absolute value below 0.09 mm/fraction for Left-Right, 0.17 mm/fraction for Anterior-Posterior and Head-Feet direction. Distribution of slope for the conventional parametrization is more narrow and has a mean value close to zero (see Fig. 4.11).

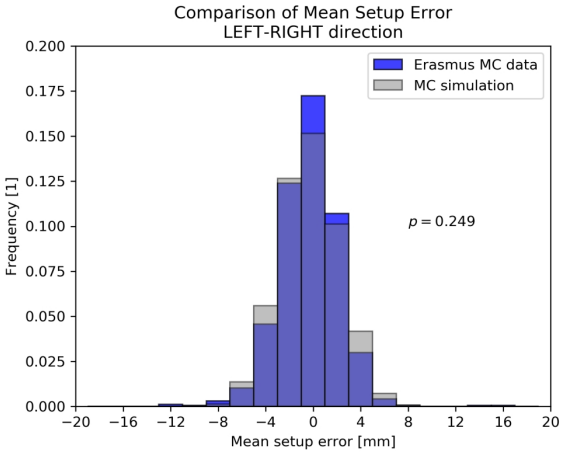
Trendline Parametrization

Fig. 4.13-4.17 present comparison of distributions for parameters used to evaluate and parametrize database of setup errors between real data and MC simulations with trendline parametrization. It can be seen that there is a significant difference between the real database and simulation in distribution of SD of random errors while using Gaussian function to randomize SD_p. Also d_{100} is not fully reproduced in that case. With the trendline parametrization the slope and offset distributions are well reproduced. Especially for the trendline parametrization with Burr distribution used for random error generation all parameters are reproduced properly in terms of statistical analysis (see Table 4.1).

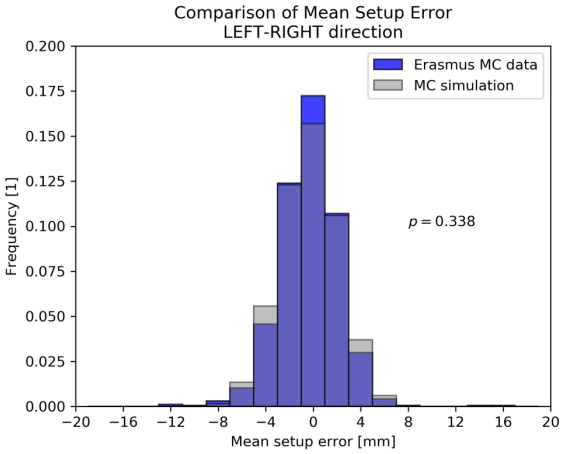
4.3.2.2 NAL and eNAL verification protocols performance

Fig. 4.18-4.20 show comparison between performance of MC parametrization and simulation methods with clinical data for NAL protocol. It can be seen that the conventional parametrization (blue lines) underestimates the residual Σ_{res} (0.7 mm) when clinically relevant number of fractions n is considered. The trendline parametrization (red lines) works better but still underestimates Σ_{res} by about 0.3 mm. The difference between using Gauss or Burr distribution to randomize SD_p is small. Usage of Burr distribution generally gives slightly smaller Σ_{res}. That can be explained by shift to smaller values of SD_p with Burr distribution compared to Gauss distribution. The smaller SD_p the better estimation of mean value is. That mean value is used as a correction factor in NAL protocol. Interestingly, NAL protocol performance for Erasmus MC data depends on order of data. An inversion gives smaller Σ_{res} and that one is well reproduced with the trendline parametrization. Σ_{res} dependence on order of fractions would be discussed in 4.4.

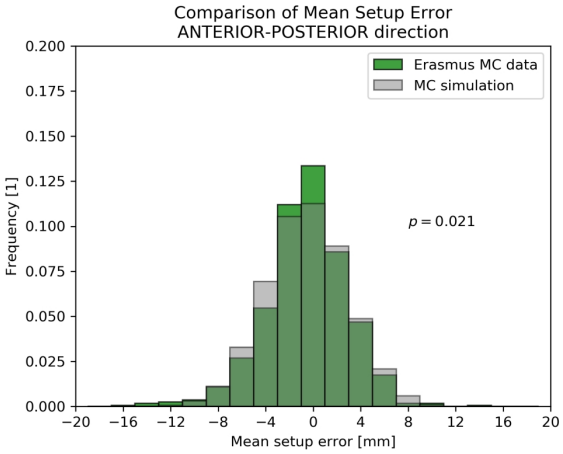
Table 4.2 presents a comparison between parametrization methods in terms of simulating NAL $N_m = 3$ and eNAL protocols performance for Erasmus MC database. The trendline parametrization results for NAL protocol are closer to direct simulation than for conventionl simulation. This is in agreement with Fig. 4.18-4.20. For eNAL both parametrizations worked in comparable way.



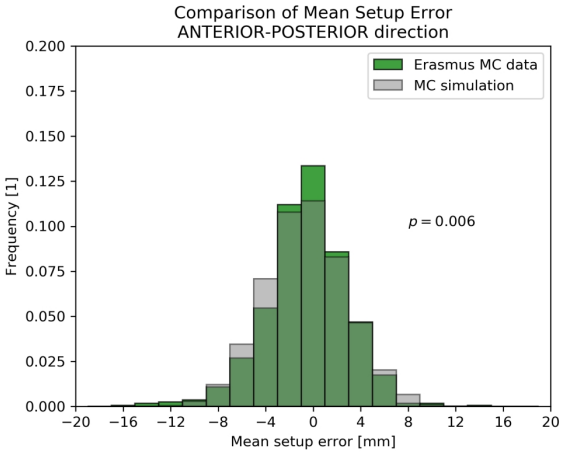
(a) Gauss distribution of random errors



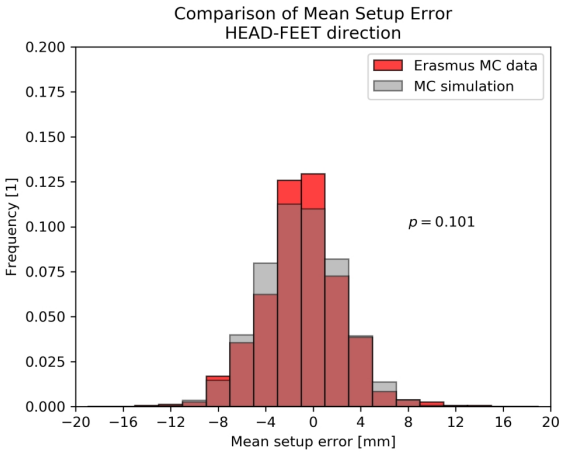
(b) Burr distribution of random errors



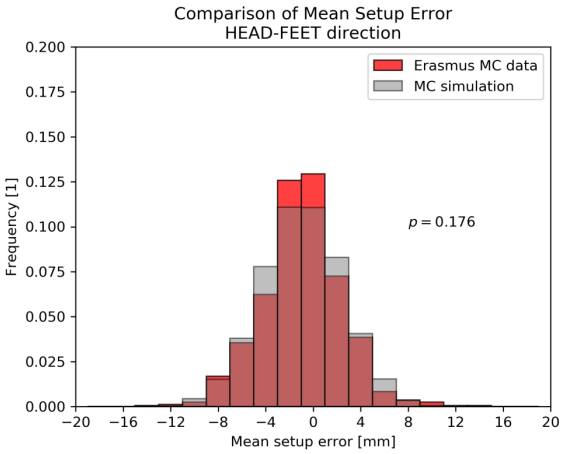
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

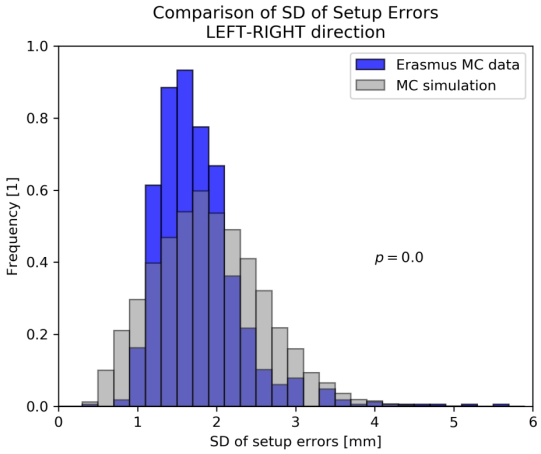


(e) Gauss distribution of random errors

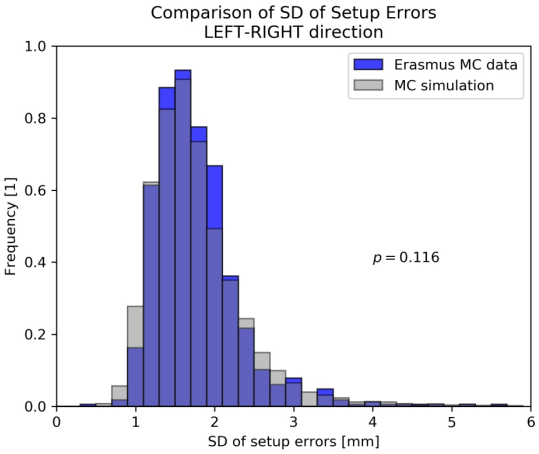


(f) Burr distribution of random errors

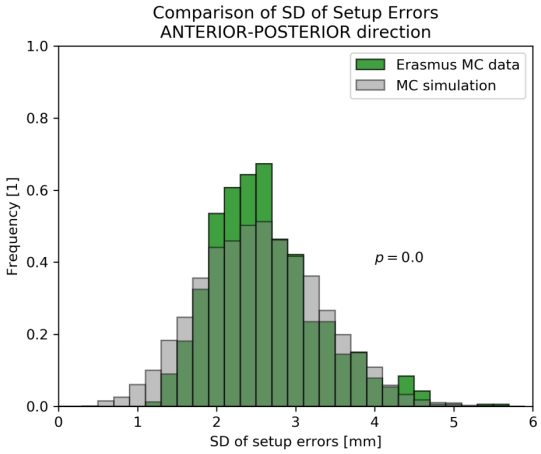
Figure 4.8: Comparison between histograms of mean setup error achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



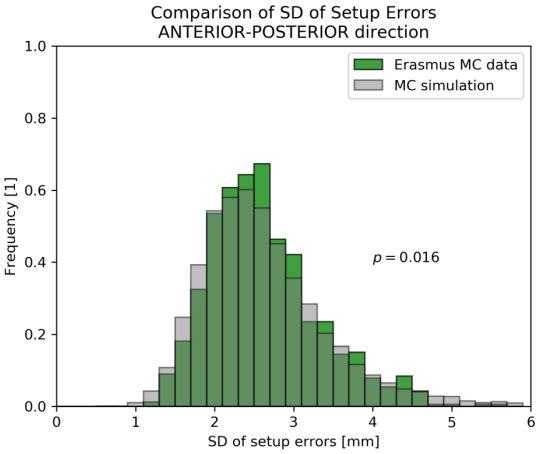
(a) Gauss distribution of random errors



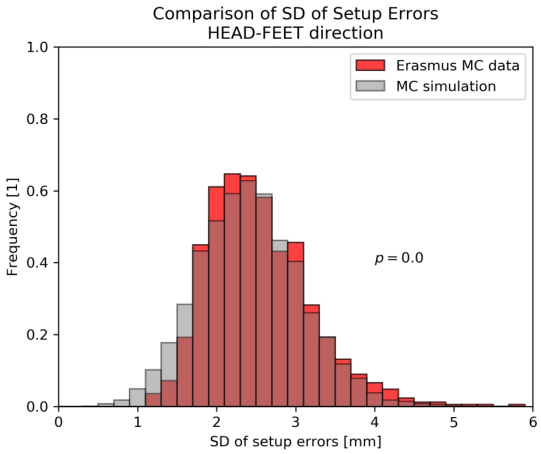
(b) Burr distribution of random errors



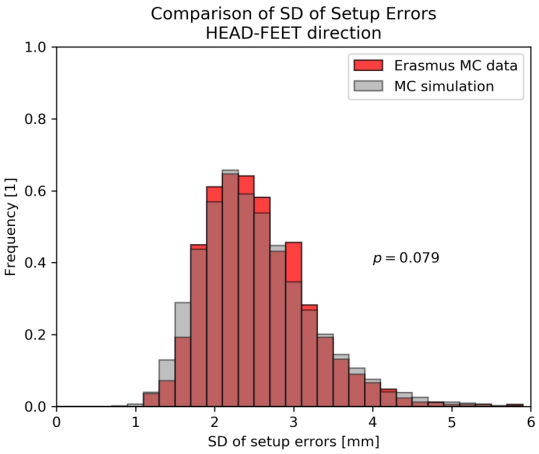
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

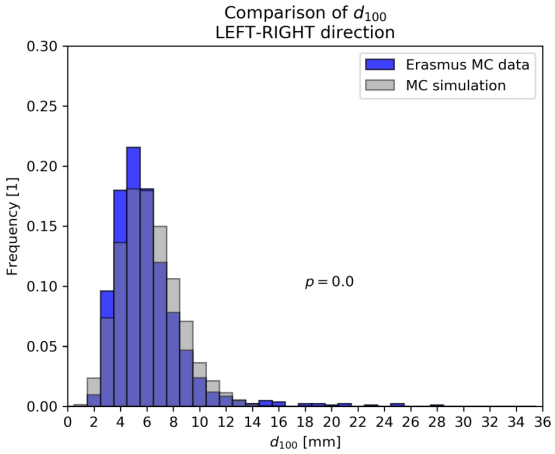


(e) Gauss distribution of random errors

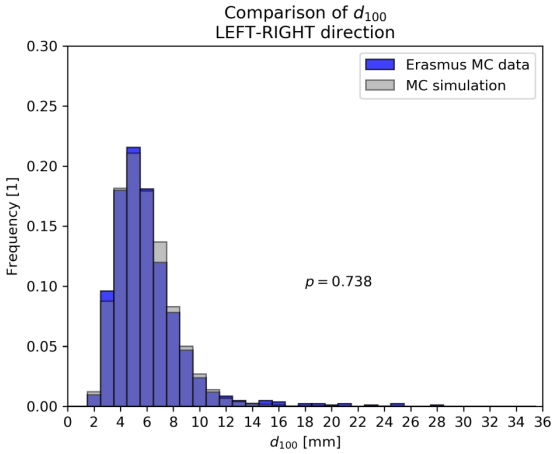


(f) Burr distribution of random errors

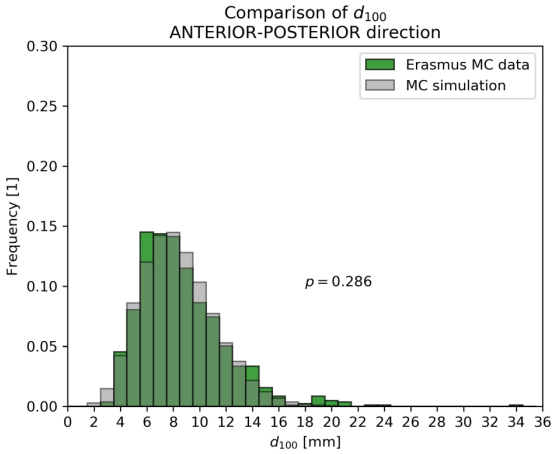
Figure 4.9: Comparison between histograms of setup errors SD achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



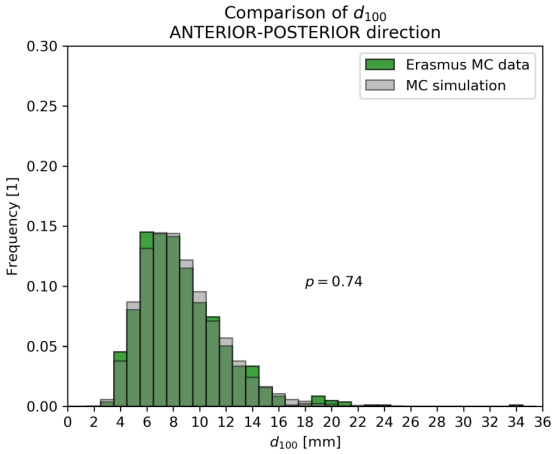
(a) Gauss distribution of random errors



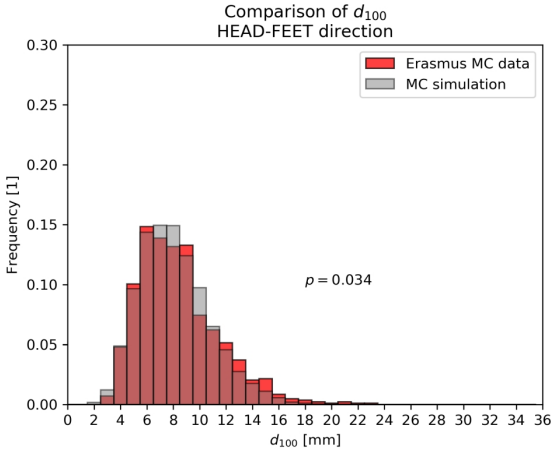
(b) Burr distribution of random errors



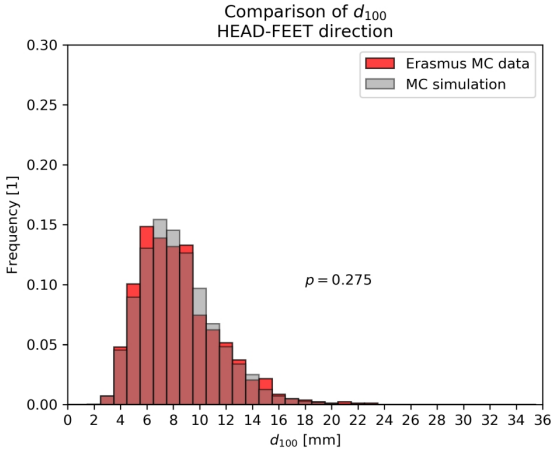
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

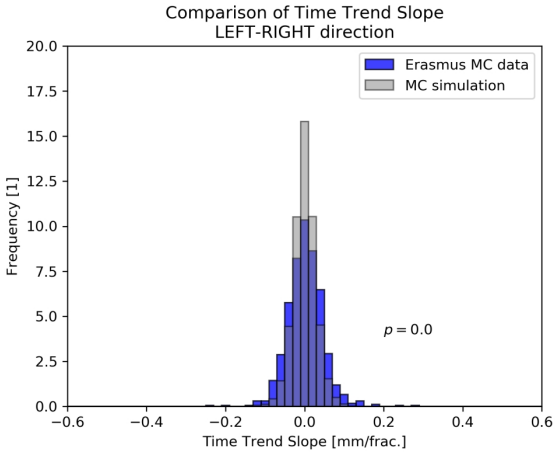


(e) Gauss distribution of random errors

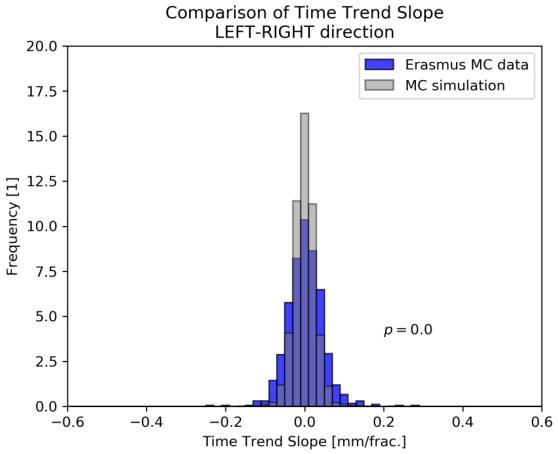


(f) Burr distribution of random errors

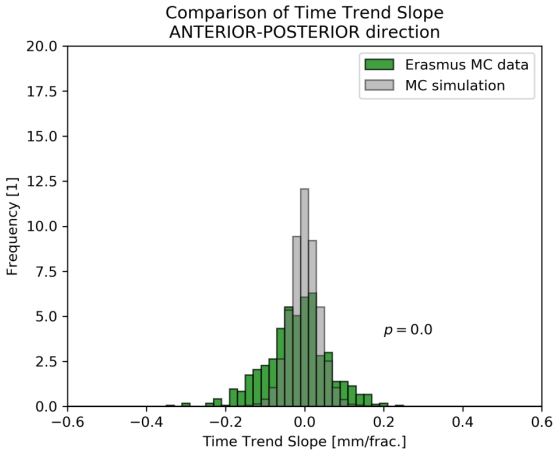
Figure 4.10: Comparison between histograms of d_{100} achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



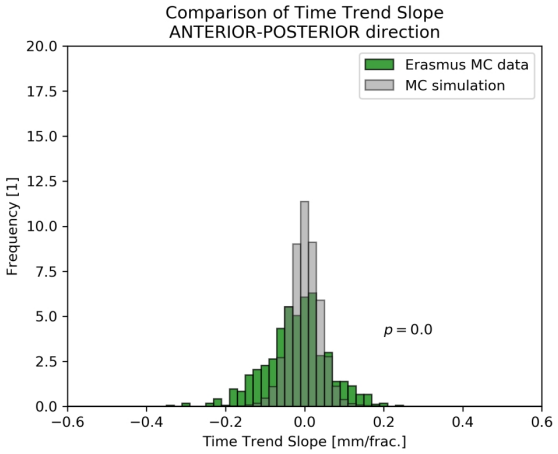
(a) Gauss distribution of random errors



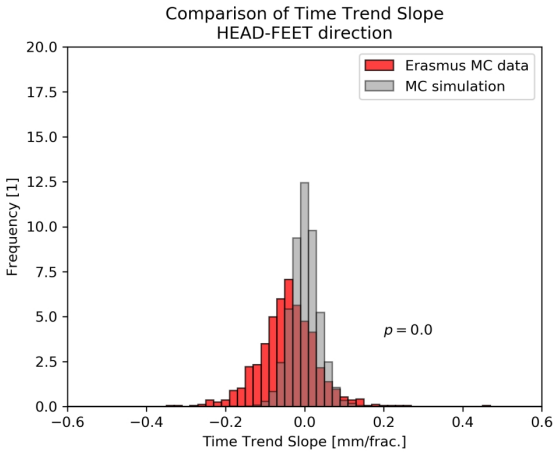
(b) Burr distribution of random errors



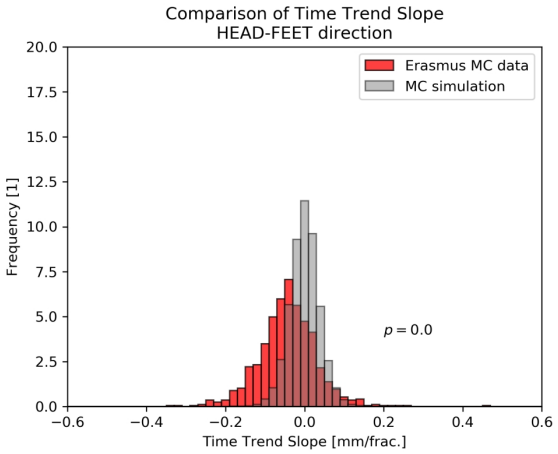
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

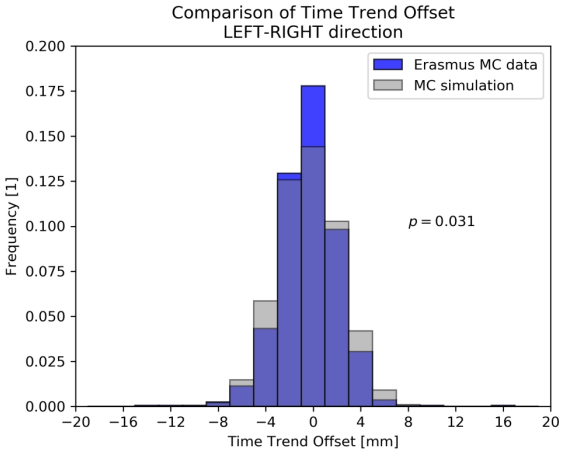


(e) Gauss distribution of random errors

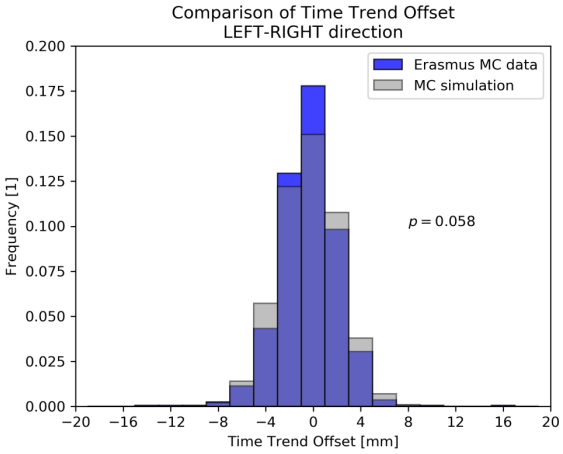


(f) Burr distribution of random errors

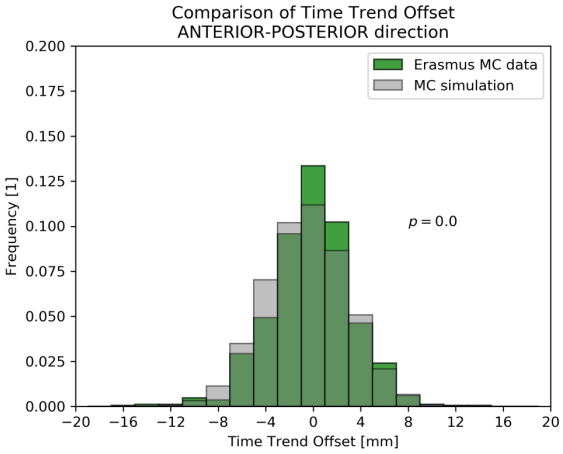
Figure 4.11: Comparison between histograms of time trend slope achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



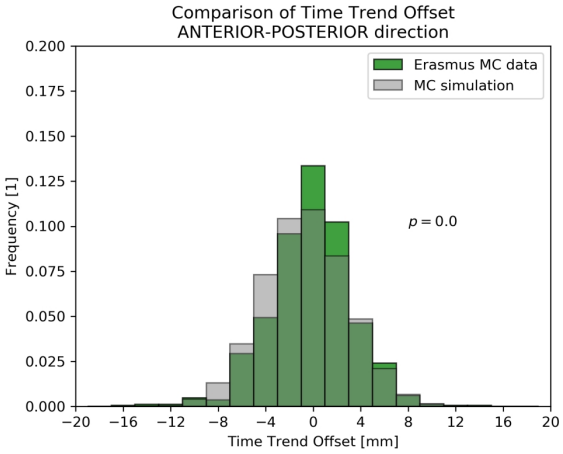
(a) Gauss distribution of random errors



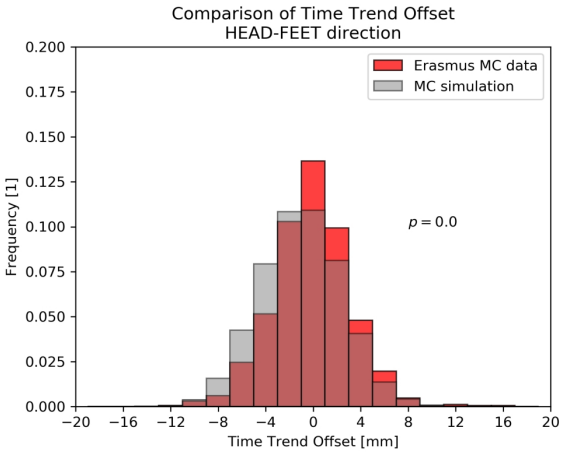
(b) Burr distribution of random errors



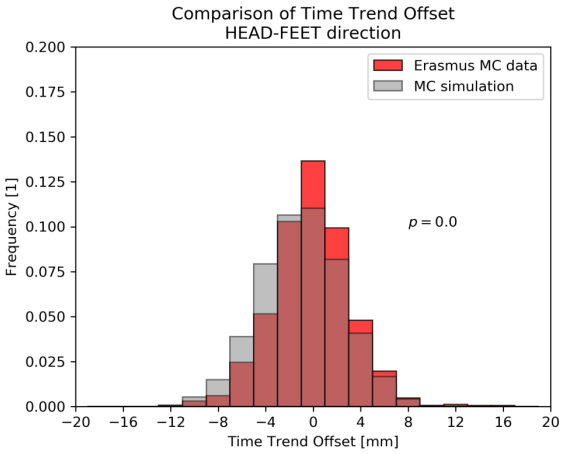
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

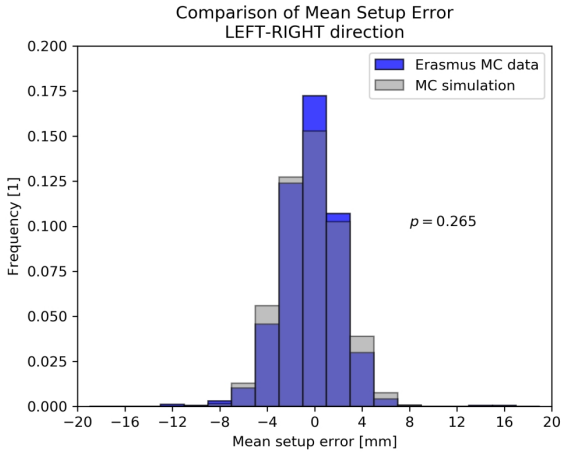


(e) Gauss distribution of random errors

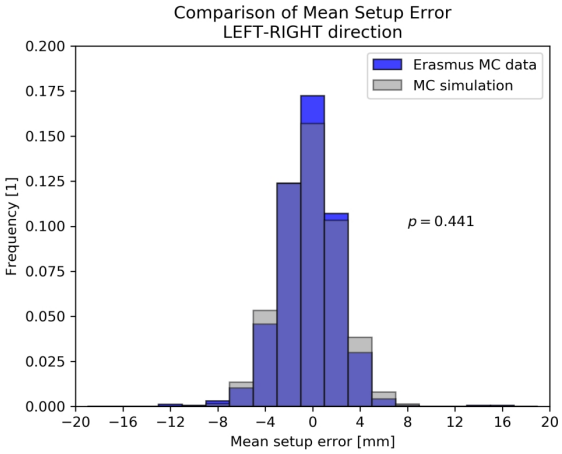


(f) Burr distribution of random errors

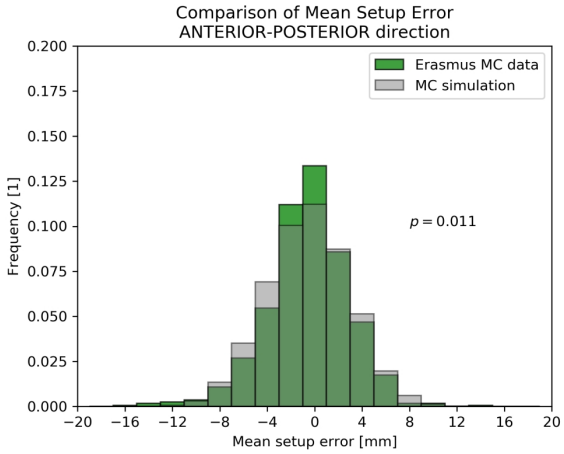
Figure 4.12: Comparison between histograms of time trend offset achieved for Erasmus MC database and MC simulations with standard parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



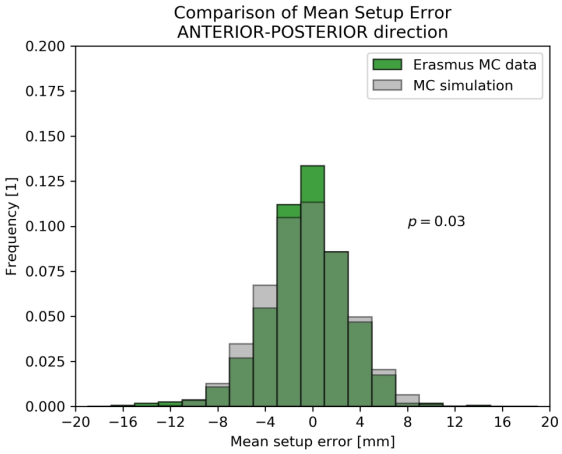
(a) Gauss distribution of random errors



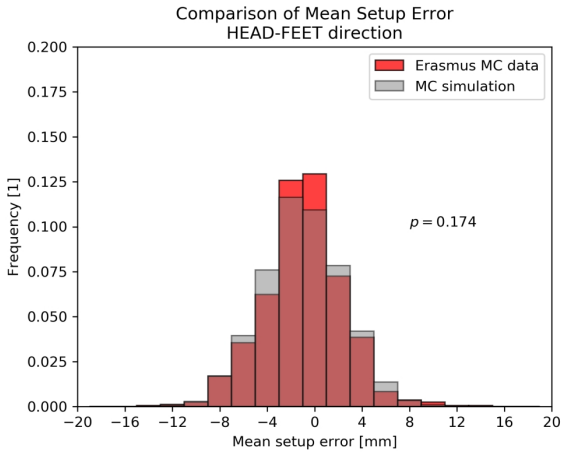
(b) Burr distribution of random errors



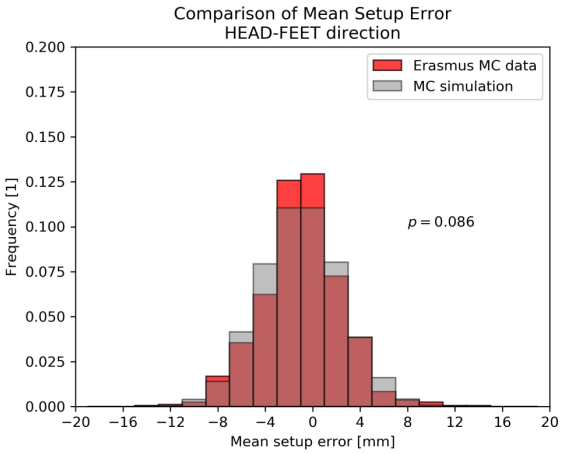
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

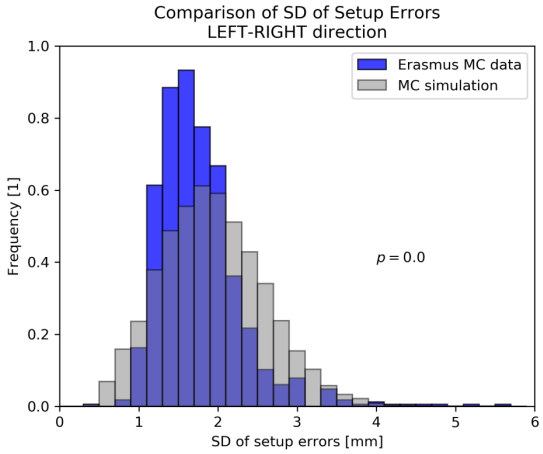


(e) Gauss distribution of random errors

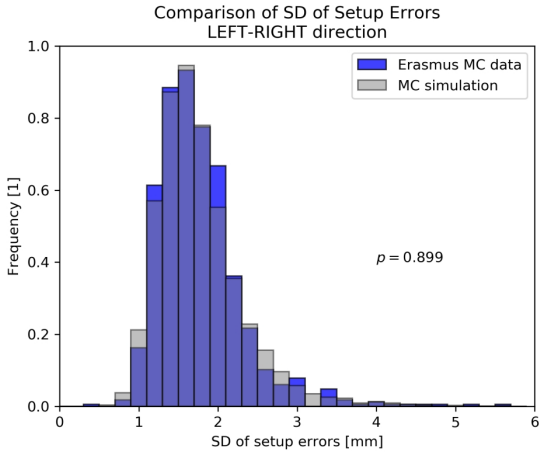


(f) Burr distribution of random errors

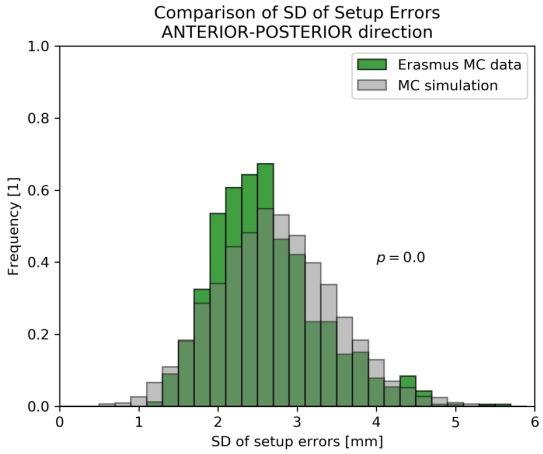
Figure 4.13: Comparison between histograms of mean setup error achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



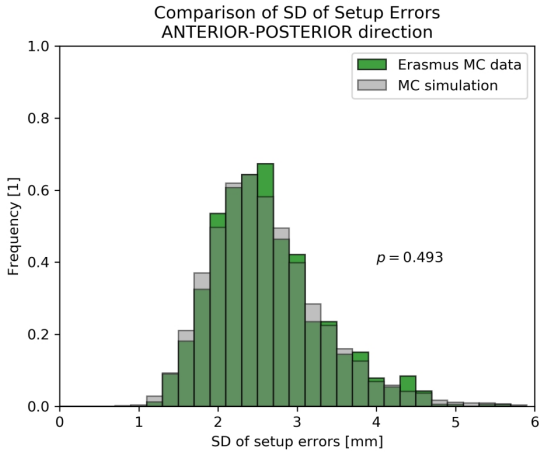
(a) Gauss distribution of random errors



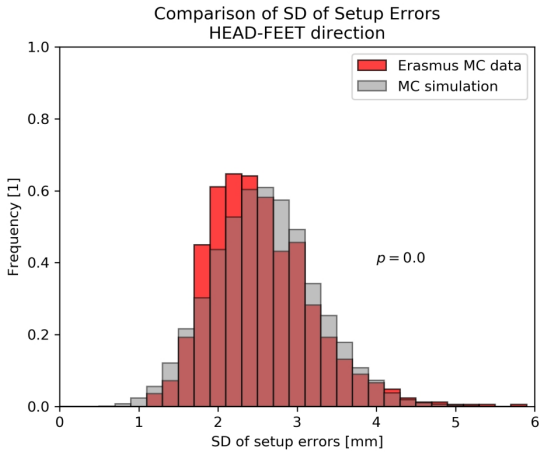
(b) Burr distribution of random errors



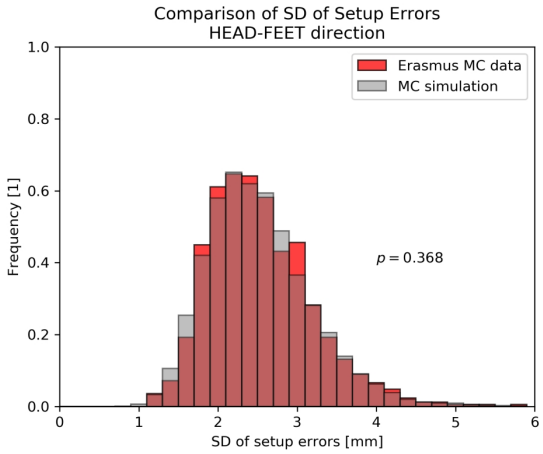
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

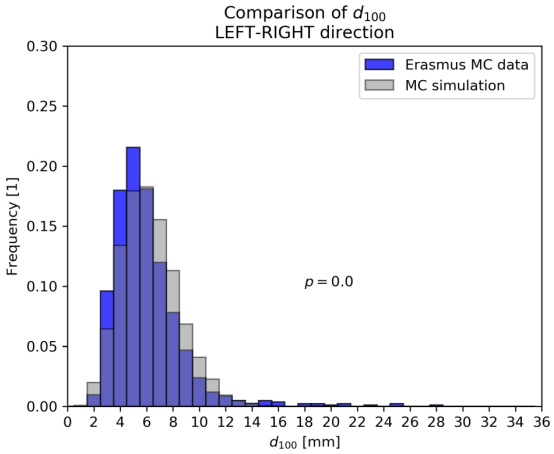


(e) Gauss distribution of random errors

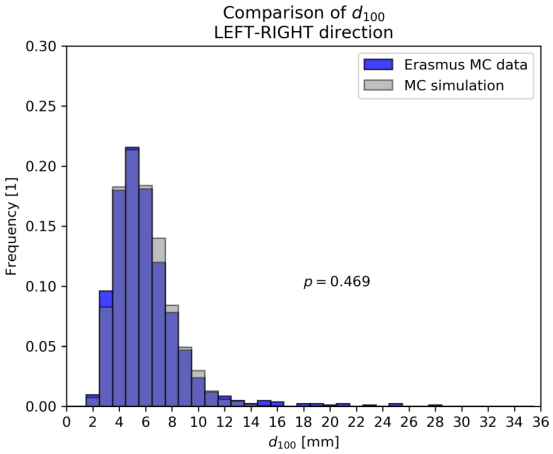


(f) Burr distribution of random errors

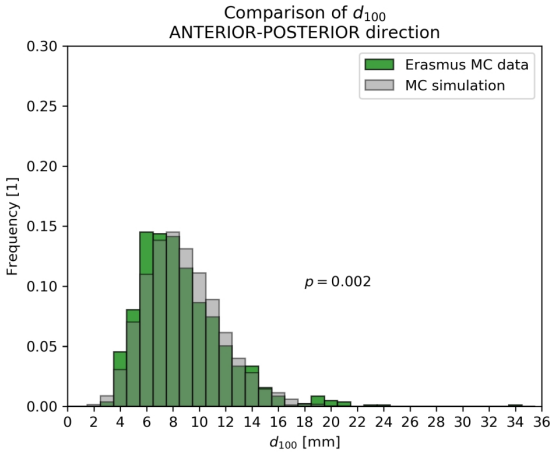
Figure 4.14: Comparison between histograms of setup errors SD achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



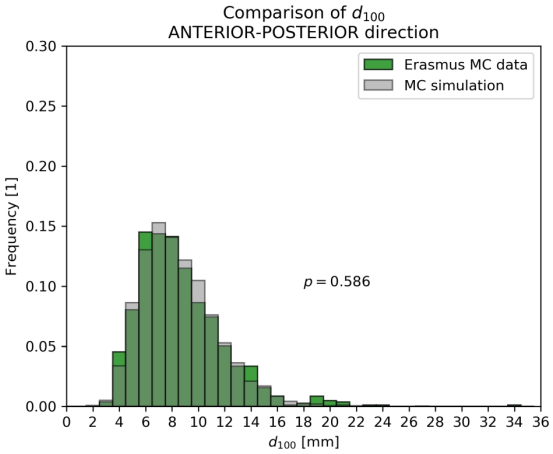
(a) Gauss distribution of random errors



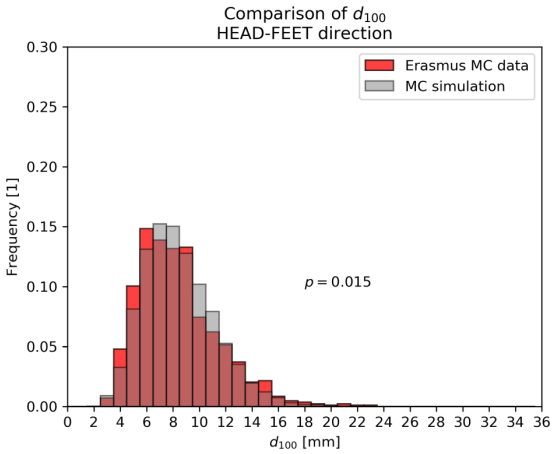
(b) Burr distribution of random errors



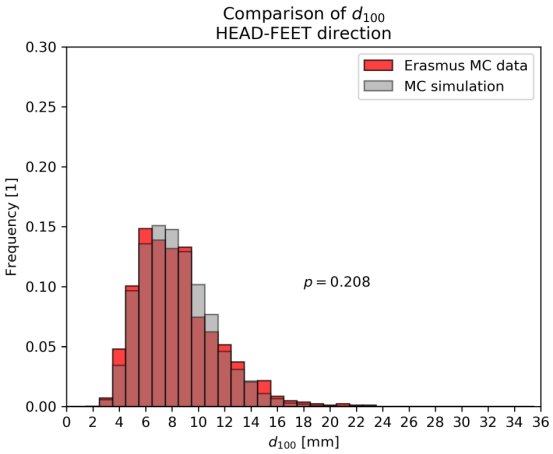
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

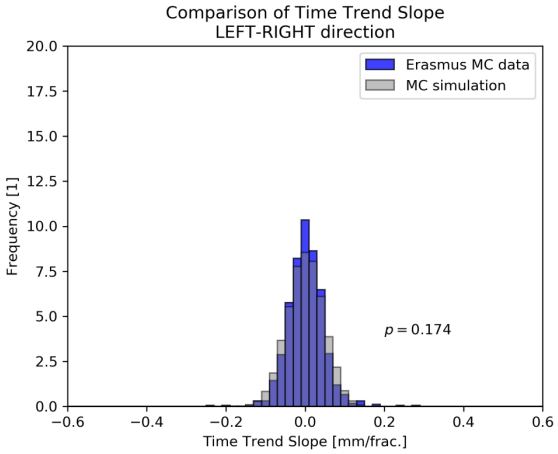


(e) Gauss distribution of random errors

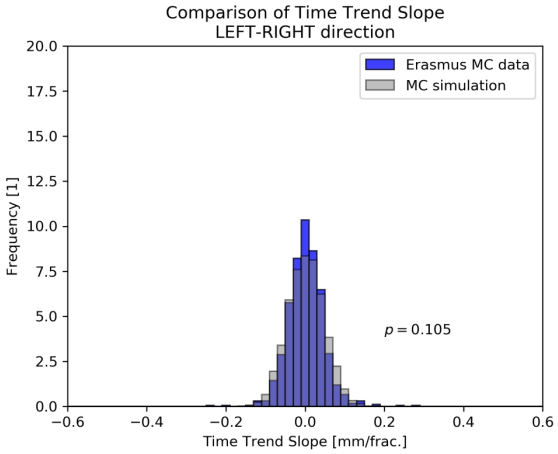


(f) Burr distribution of random errors

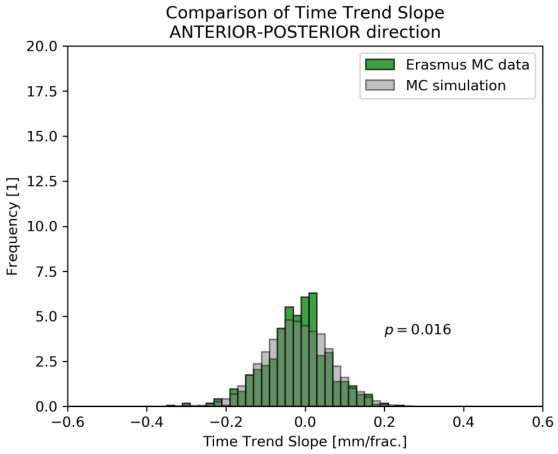
Figure 4.15: Comparison between histograms of d_{100} achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



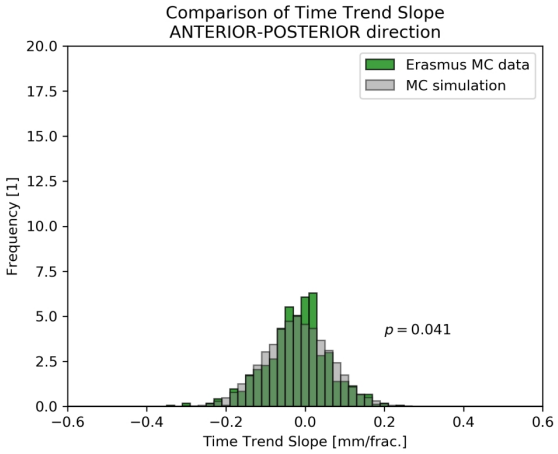
(a) Gauss distribution of random errors



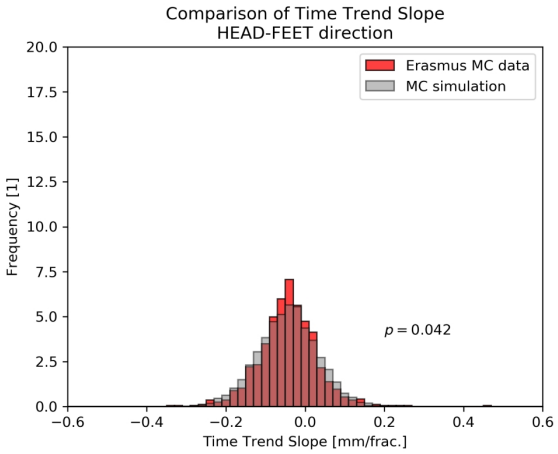
(b) Burr distribution of random errors



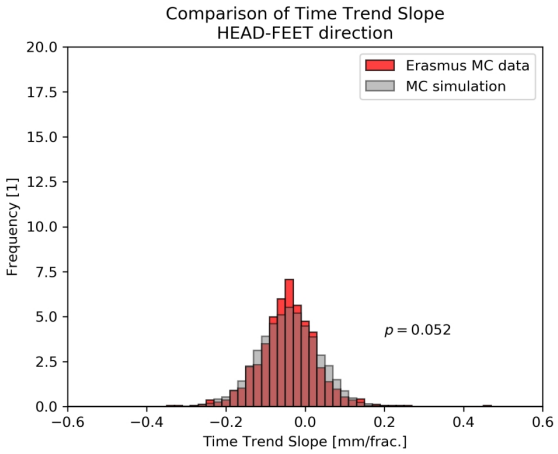
(c) Gauss distribution of random errors



(d) Burr distribution of random errors

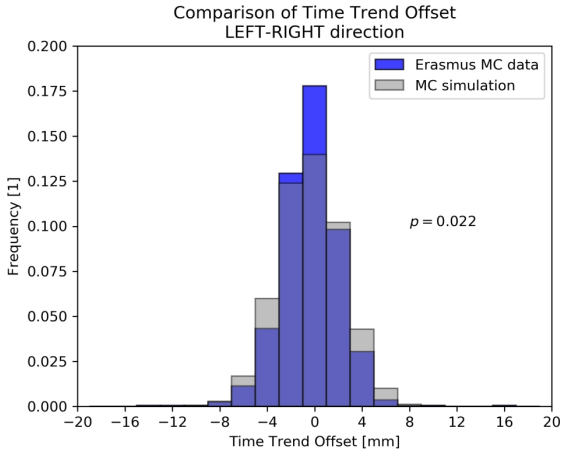


(e) Gauss distribution of random errors

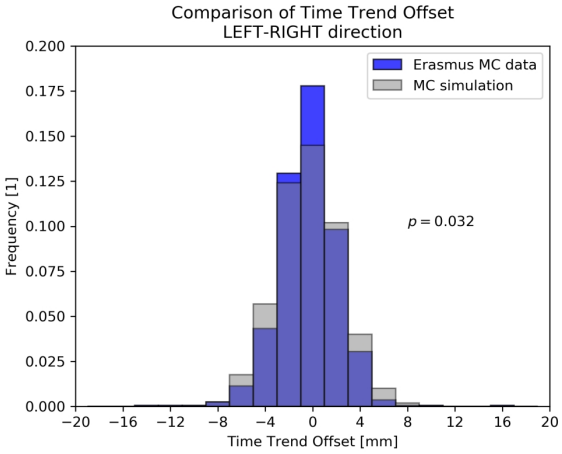


(f) Burr distribution of random errors

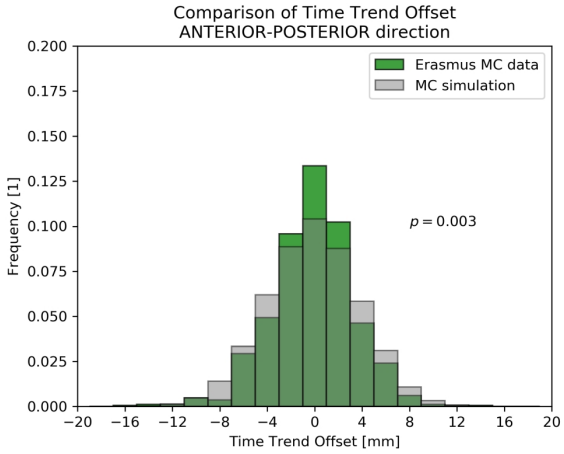
Figure 4.16: Comparison between histograms of time trend slope achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.



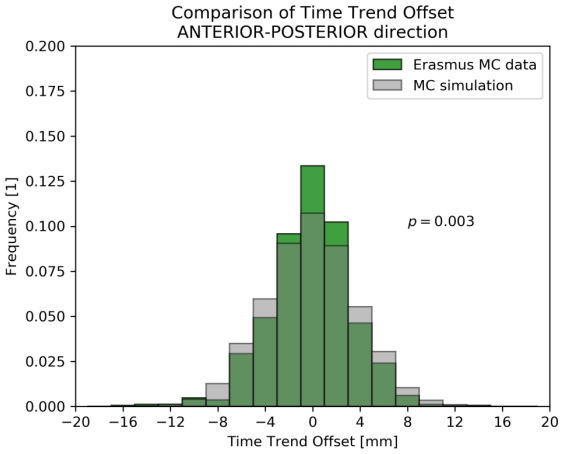
(a) Gauss distribution of random errors



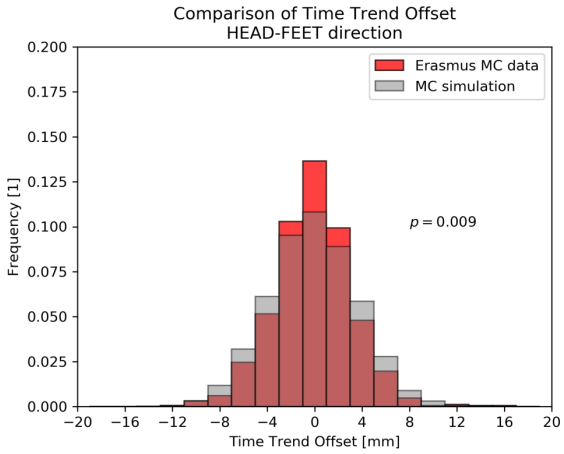
(b) Burr distribution of random errors



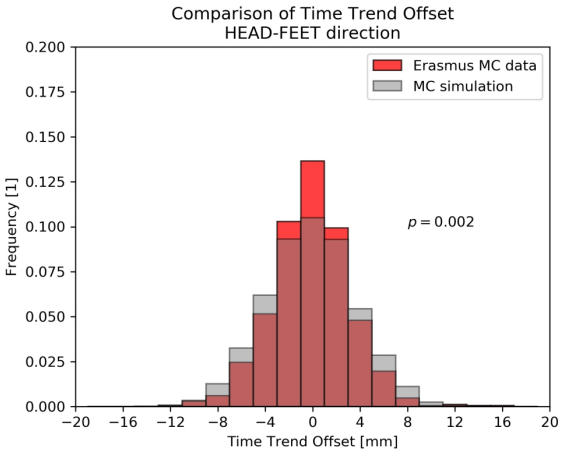
(c) Gauss distribution of random errors



(d) Burr distribution of random errors



(e) Gauss distribution of random errors



(f) Burr distribution of random errors

Figure 4.17: Comparison between histograms of time trend offset achieved for Erasmus MC database and MC simulations with trendline parametrization. On the left the comparison is done for MC simulations assuming Gaussian distribution of random error, on the right — Burr distribution was assumed.

	NAL			eNAL		
	Direct simulation	Trendline MC	Conventional MC	Direct simulation	Trendline MC	Conventional MC
LR	1.4	1.2	1.1	0.8	1.1	1.0
AP	2.0	1.7	1.3	0.8	1.1	1.0
HF	2.1	1.9	1.4	0.8	1.1	1.0

Table 4.2: Table presents Σ_{res} values achieved for Erasmus MC database in left-right (LR), anterior-posterior (AP) and head-feet (HF) direction after the NAL ($N_m = 3$) and eNAL protocols were simulated. All presented values are given in mm.

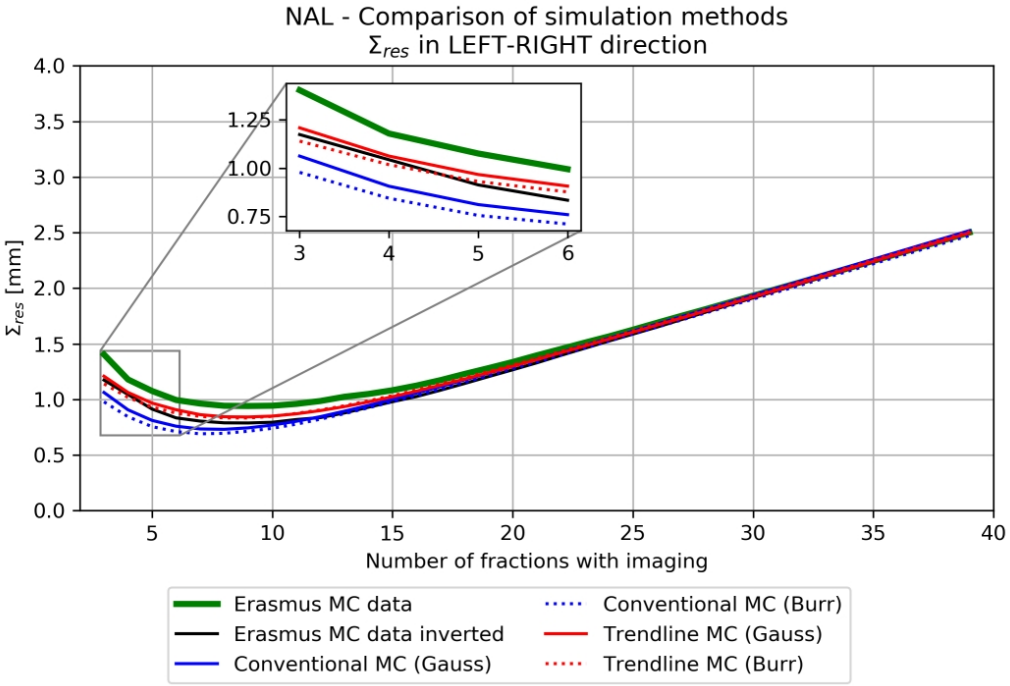


Figure 4.18: Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in left-right direction.

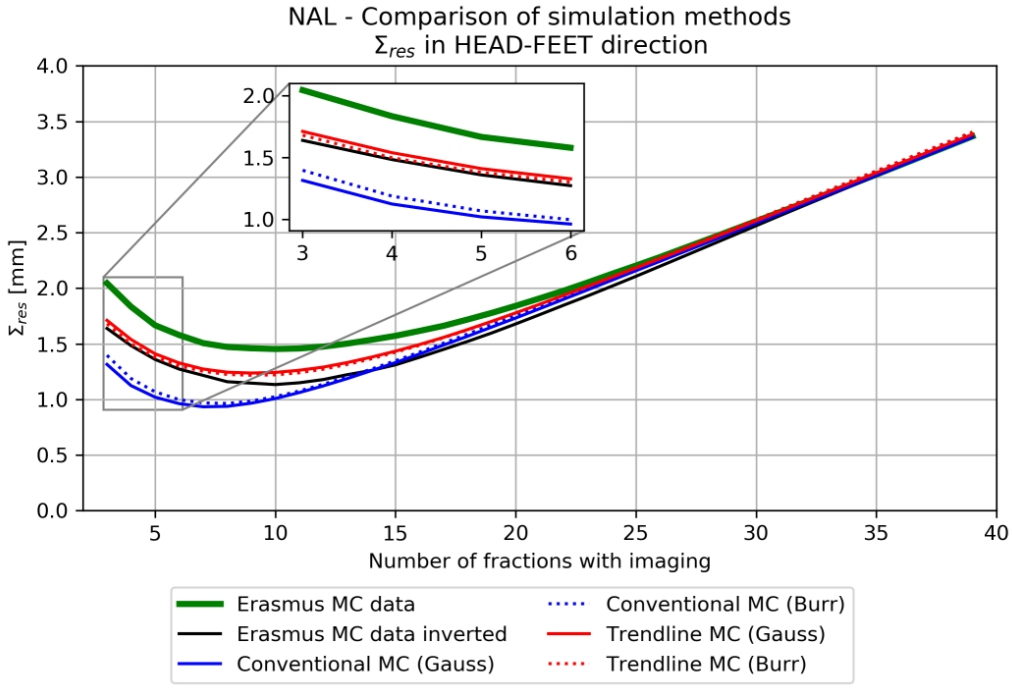


Figure 4.19: Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in head-feet direction.

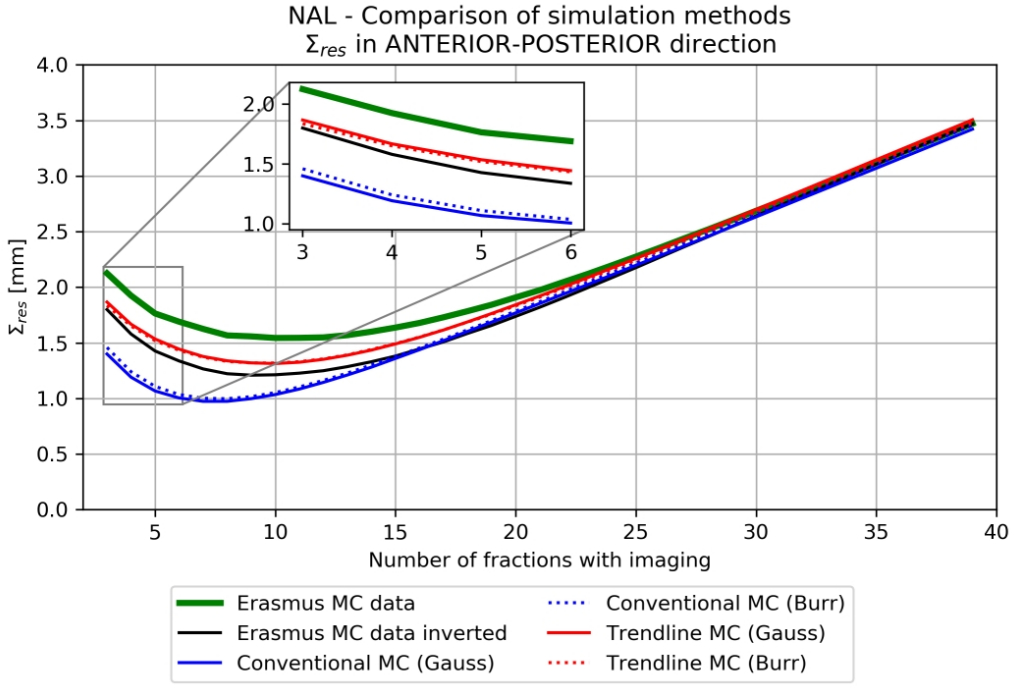


Figure 4.20: Comparison of Σ_{res} achieved with NAL protocol for Erasmus MC database and MC populations in anterior-posterior direction.

4.4 Discussion

Conventional parametrization of patient population setup errors was previously used to test efficiency of OVP [7, 14]. Parametrization incorporating time trends was previously described by de Boer and Heijmen [16]. In this chapter this parametrization was further explored and widely compared with the conventional parametrization. According to findings presented in 4.3 if significant time trends are observed the conventional parametrization does not give the right model. The trendline parametrization should be used instead, otherwise Σ_{res} can be 50%-70% underestimated and in consequence too small CTV-PTV margin might be applied. That is why it is of crucial importance to test setup errors in patients population for existence of time trends. The method how to do that was proposed in Section 3.6.

In Section 4.3.2.2 it was shown that the trendline parametrization does not reproduce NAL performance for Erasmus MC database (although it is closer to ground truth than conventional parametrization). Σ_{res} strongly depends on goodness of estimation of the mean setup error $\overline{m_p}$. The greater is SD_p the worse is the estimation. Usually an assumption is made that SD_p is constant during the radiotherapy course. That was also assumed in provided MC simulations. In a case the assumption is true if we reverse order of fractions the result should be the same. It was the case in synthetic populations (see Fig. 4.21). In case of Erasmus MC database inversion of fraction order gives different results (see Fig. 4.18-4.20). It can be seen that the reverse order of fractions gives similar results for the simulation with trendline parametrization. That indicates that SD'_p taken in simulation was of the order of last part of the treatment. It has to be kept in mind that in fact SD'_p was calculated for whole treatment. From the difference between normal and reversed order of fractions it can be presumed that SD_p (and SD'_p) are changing during time of treatment. In order to check that hypothesis the change of σ in time was evaluated. Eq. (2.7) was used to calculate σ for each following 3 fractions. In order to do that setup errors for following 3 fractions (fractions 1-3, 2-4, ..., 37-39) were extracted from the database. Due to the fact that 3 is a small number of fractions the correction for unbiased estimation was done (σ was divided by $\sqrt{\pi}/2$). The results are presented in Fig. 4.22. Especially for anterior-posterior direction it can be seen that σ is decreasing in time. That trend seen for population does not have to be the case for a particular patient.

Patient setup errors and especially random setup errors can be influenced by many things, like: stress, family life, waiting time, treatment staff, diet, being familiar with equipment and so on. During the course of radiotherapy patient should be more and more familiar with the treatment machine and radiotherapy process — that should generally decrease the SD_p of random error. Although it might happen that at the end of therapy patient would be bored and nervous especially if it cost him a lot of effort to be on time and he has to wait for irradiation afterwards. Everyday life with its stress, troubles, hope and joy is highly unpredictable. That is why I think that stability of SD_p should not be assumed. More other further methods of investigating SD_p change in time during course of treatment should be applied.

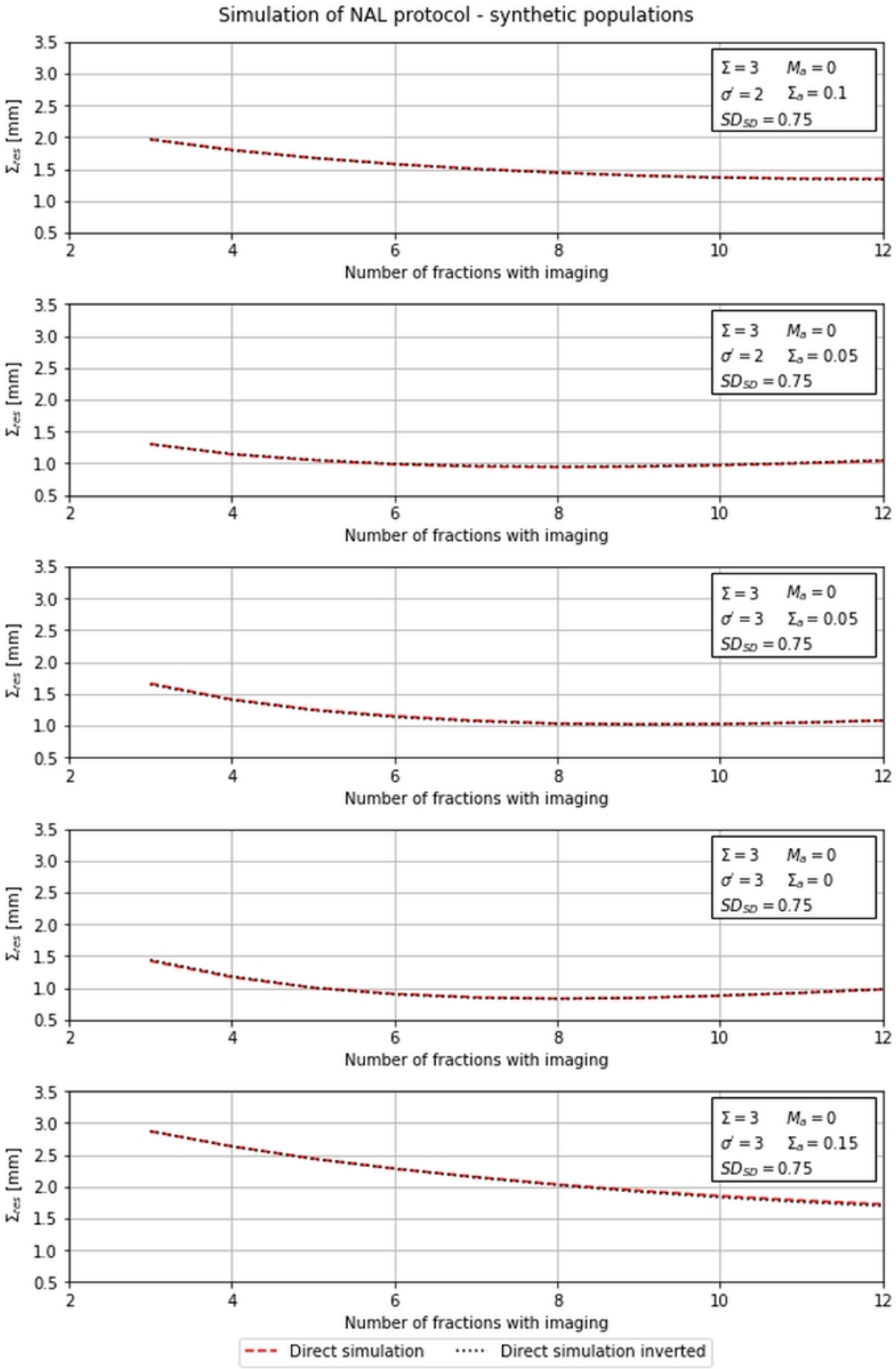


Figure 4.21: Comparison of Σ_{res} achieved with NAL protocol for arbitrary chosen synthetic populations (direct simulation). Performance in case of inverted fraction order is also shown

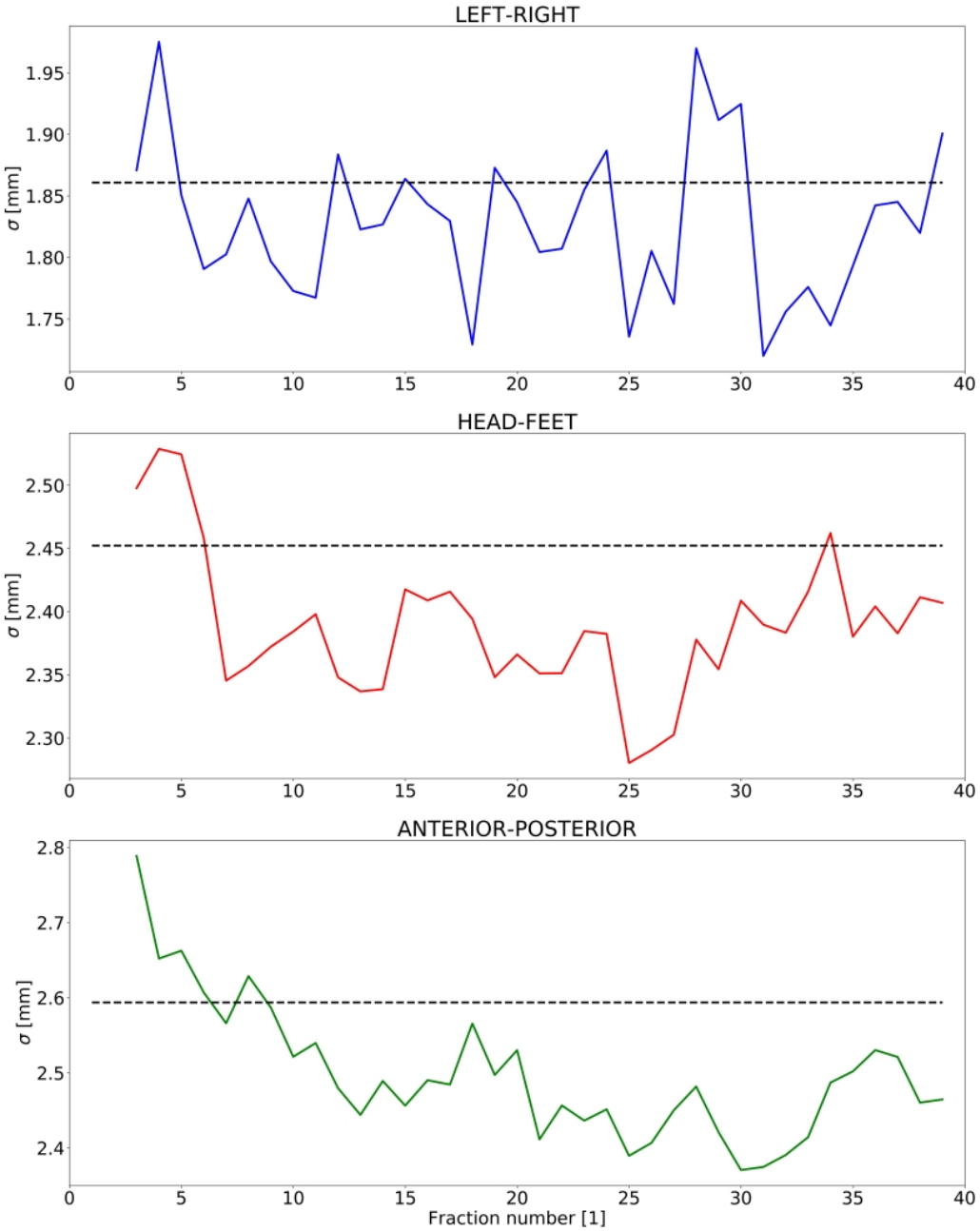


Figure 4.22: Examination of $\sigma/0.5\sqrt{\pi}$ changes in time.

Chapter 5

Estimation of Cumulative Dose

5.1 Background

CTV-PTV margins designed on the basis of van Herk recipe (see Section 2.2) are population based ones. The individual patient parameters ($\overline{m_p}$, SD_p - see Eq. 2.3-2.4) are not taken into account during creation of these margins. Moreover, their definition assume the normal distribution of setup errors which is not always the case (see Chapter 3). That is why I decided to check whether it would be possible to predict, in other words — to estimate, the cumulative dose distribution for an individual patient after few first fractions of his treatment. If so, methods of margin individualization might be considered. A treatment plan individualization in terms of setup errors can influence dose both in PTV and OARs. Individualized margins may give smaller field sizes and consequently decrease irradiated volume of healthy tissue. Much effort is done nowadays to perform individualization in PTV definition [46]. Some authors propose to include uncertainties into the optimization step in the margin-less planning concept [5, 27, 62, 63], others propose the concept of probability PTV [85]. Both methodologies require changes in TPS and therefore have to be implemented by vendors into new algorithms. For clinical users it means the need of buying a new software and/or license. The simple method of cumulative dose estimation proposed and verified in this work does not require any change in TPS. Therefore it can be used with current software still allowing individualization of patient treatment.

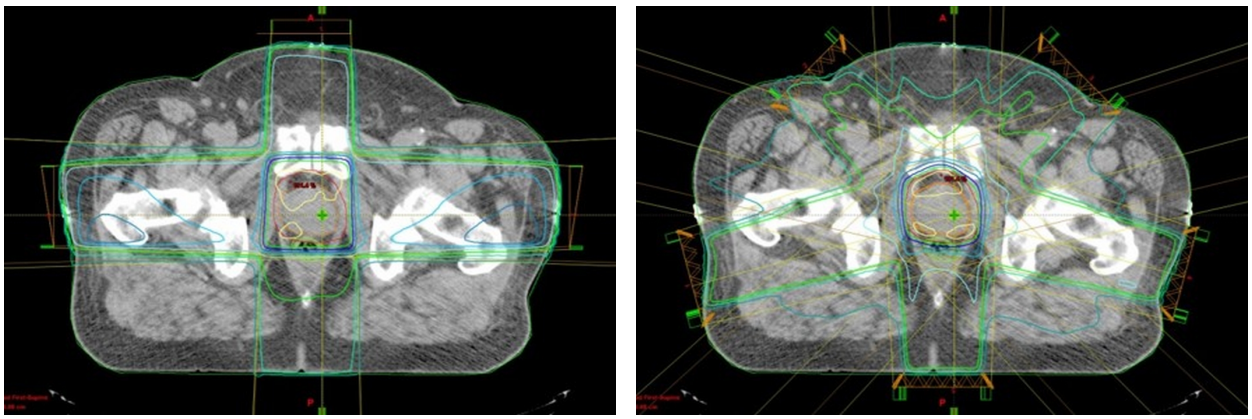
5.2 Material and methods

5.2.1 Clinical Data

Analysis of estimation of the cumulative dose distribution was done retrospectively for 25 patients who underwent 3D-CRT radiotherapy in COI. Those patients were a subgroup from the patients described in Section 3.2.1. For all patients setup errors in all 25 fractions were recorded during their treatment.

5.2.2 Radiotherapy Treatment Plans

All plans described below were prepared retrospectively. Eclipse 10.0 (Varian Medical Systems Inc., Palo Alto, USA) TPS with Analytical Anisotropic Algorithm (AAA) version 10.0.28 was



(a) 3D-CRT (b) IMRT

Figure 5.1: Treatment plan geometry for 3D-CRT (Fig. 5.1a) and IMRT (Fig. 5.1b) plans.

used. For each patient plans in three different techniques (3D-CRT, IMRT, VMAT) were prepared by an experienced planner. Prescribed dose in each plan was 65 Gy in 25 fractions.

5.2.2.1 3D-CRT

3D-CRT plans consist of two lateral fields (90°and 270°) and one anterior beam: 0°(see Fig. 5.1a). All beams were 15 MV. Both lateral beams had wedges. Multileaf Collimator (MLC) position was set automatically with 0.9 cm distance to PTV in head-feet direction and 0.7 cm distance in other directions.

5.2.2.2 IMRT

IMRT plans consist of five 15 MV fields with gantry angle of: 255°, 315°, 45°, 105°and 180°(see Fig. 5.1a). MLC delivery was Sliding Window. Varian Leaf Motion Calculator was used to calculate MLC pattern on the basis of optimal fluence, which is the results of optimization. Inverse planning was done interactively in Helios (dedicated optimizer implemented in Eclipse). Optimization parameters were changed during optimization process by experienced treatment planner in order to achieve plan reaching such criteria as: $D_{99\%}^{PTV} >= 95\%$, $V_{61\text{ Gy}}^{rectum} <= 25\%$, $V_{52\text{ Gy}}^{bladder} <= 50\%$.

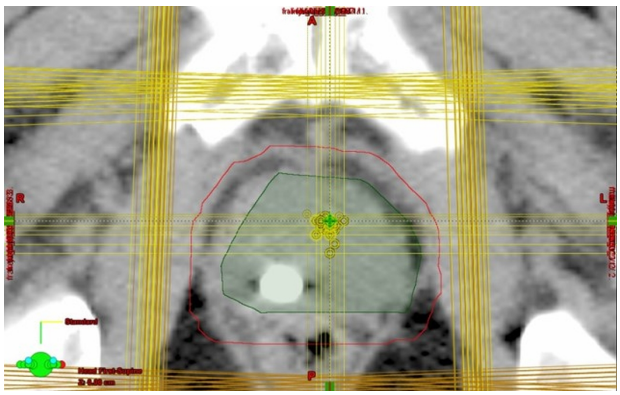
5.2.2.3 VMAT

VMAT plans (also called RapidArc plans) consist of two full, coplanar, 6 MV arcs. The MLC pattern, dose rate and gantry speed were optimized with the same criteria as for IMRT.

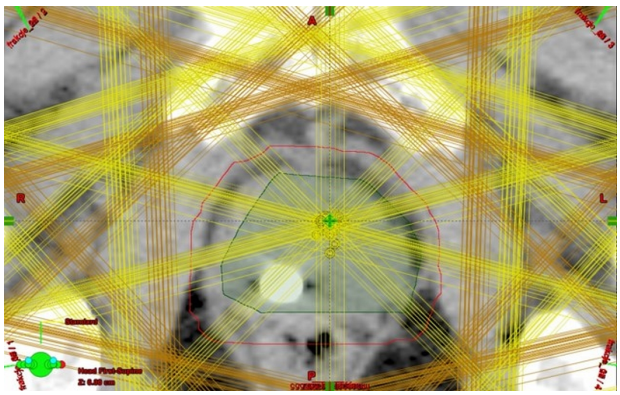
5.2.3 Cumulative Dose Distribution

Cumulative dose distribution D_{cum} is defined as the sum of dose distributions calculated for all treatment fractions separately. Here it is the sum of dose distributions in case no setup corrections were applied.

Gathered setup error data consisted of $\boldsymbol{m_{p,f}}$. Setup errors were measured in relation to bony anatomy. In order to implement this setup errors into the TPS the isocenter position



(a) 3D-CRT



(b) IMRT

Figure 5.2: Distribution of isocenter positions for 3D-CRT (Fig. 5.2a) and IMRT (Fig. 5.2b) plans.

had been shifted by $-\boldsymbol{m_{p,f}}$. Dose distributions were recalculated for the shifted isocenter with preserved Monitor Units (MU) for each field. MU values were rounded to machine precision. The dose distribution calculated with isocenter shifted for particular setup error was considered to be the best representative of dose distribution in specified fraction (see Fig. 5.2). It is not ideal dose distribution due to the intra- and inter- fraction changes in patient anatomy. The only CT used in this analysis was that done before RT planning.

It has to be emphasized that it is known that in case of prostate irradiation prostate changes its position depending on bladder and rectum filling [13]. Therefore in many hospitals, also in COI, CBCT is used in order to minimize setup error. Still there are other treatment sites for which tumor does not move relative to body anatomy. It is also important to mention that for prostate cancer patients, in COI, setup errors are gathered in each fraction during normal treatment (protocol described in Section 2.3.1). That is why I decided to test methods of estimating cumulative dose distribution for data that were already gathered [33].

Denoting $D_{p,f}(\boldsymbol{m_{p,f}})$ as the dose distribution in fraction f for patient p calculated with isocenter shifted by $-\boldsymbol{m_{p,f}}$, cumulative dose for patient p — $D_{p,cum}$ can be derived as:

$$D_{p,cum} = \sum_{f=1}^{n_f} D_{p,f}(\boldsymbol{m_{p,f}}) \tag{5.1}$$

5.2.4 Methods of estimating Cumulative Dose Distribution

Cumulative dose distribution can be calculated only after the whole treatment and only in case all setup errors were recorded and are known. That is why methods of estimating cumulative dose distribution are searched.

5.2.4.1 Estimation with mean setup error

The first idea of estimating $D_{p,cum}$ was to calculate the mean setup error (i.e. mean shift) — $\overline{\boldsymbol{m_{p,f}}}(n)$ other n first treatment fractions, where n would be $\{2, 3, 4, ..., 14, 15, 20, 25\}$. Afterwards the isocenter was shifted according to $\overline{\boldsymbol{m_{p,f}}}(n)$ and Dose Distribution (for n known

fractions) $D_{p,ms}(n)$ was calculated:

$$D_{p,ms}(n) = D_p(\overline{\boldsymbol{m_{p,f}}}(n)) = D_p\left(\sum_{f=1}^{n} \frac{\boldsymbol{m_{p,f}}}{n}\right) \tag{5.2}$$

5.2.4.2 Estimation with mean dose distribution

The second idea of estimating $D_{p,cum}$ was to assume that the set of setup errors from first n fractions is the best representative of the whole population of setup errors for that particular patient. The dose distribution from this first n fractions was averaged and multiplied by total number of fractions in order to calculate $D_{p,md}(n)$.

$$D_{p,md}(n) = F \cdot \frac{1}{n} \sum_{f=1}^{n} D_{p,f}(\boldsymbol{m_{p,f}}) \tag{5.3}$$

5.2.5 Gamma Evaluation

In Quality Assurance (QA) of the radiotherapy process there is often a need to compare measured and calculated dose distribution. It happens for example during TPS commissioning [60] or during pre-treatment verification of IMRT or VMAT plans [59]. Such a comparison consist of comparing two dose distributions. Dimension in space of compared dose distributions can be 1D (profile), 2D (plane) or 3D (volume). Another dimension is the dose value. In 1998 Low et al. [53] proposed a gamma index concept as a simultaneous measure of dose and distance difference between two dose distributions.

Gamma evaluation gives a single value which combines difference in dose and physical distance, both scaled with the acceptance criteria. A schematic illustration of quantities used in Eq. (5.4) is shown in Fig. 5.3. During calculation of gamma index we have to search over all points in the reference dose distribution for such a point from the evaluated dose distribution that its scaled distance in dose and physical space would be the smallest one. In order to do that we have to compare the physical distance between points in the reference dose distribution $\boldsymbol{r_r}$ and the compared dose distribution $\boldsymbol{r_c}$ and scale it by the accepted distance criteria Δd_M. In the same manner we have to compare a dose in reference dose distribution — $D_r(\boldsymbol{r_r})$ with a dose in compared dose distribution and scale it by the accepted dose difference ΔD_M.

$$\gamma(\boldsymbol{r_r}) = \min\left\{\sqrt{\frac{(D_c(\boldsymbol{r_c}) - D_r(\boldsymbol{r_r}))^2}{\Delta {D^2}_M} + \frac{|\boldsymbol{r_r} - \boldsymbol{r_c}|^2}{\Delta d_M^2}}\right\} \forall\{\boldsymbol{r_c}\} \tag{5.4}$$

Usually the Δd_M is chosen to be 2 mm or 3 mm, while ΔD_M is 2% or 3%. As ΔD_M is given as a percent value it has to be explained that it can be a fraction of maximum dose in distribution or local dose (i.e. $D_r(\boldsymbol{r_r})$). Terms of global and local gamma evaluation are used in order to differentiate between the two.

Gamma evaluation was used in order to compare cumulated and estimated dose distributions. Comparison was done with a Python script. Gamma evaluation parameters were set to 2 mm, 2%. Global as well as local gamma analysis was performed. The dose threshold was set to 10%. That means that gamma evaluation was done only for points with dose equal or larger than 10% of prescribed dose.

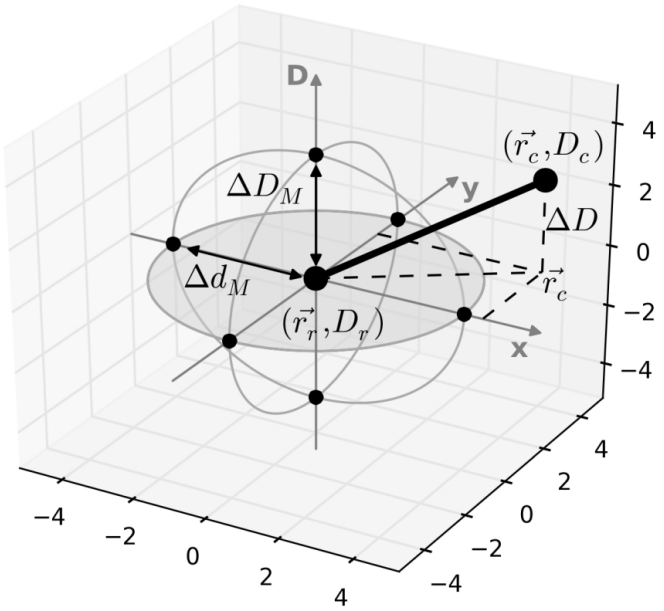


Figure 5.3: Schematic illustration of the concept of gamma evaluation method. For readability distance space is set to 2D. Point $(\boldsymbol{r_r}, D_r)$ stands for point in reference dose distribution. Point $(\boldsymbol{r_c}, D_c)$ stands for exemplary point from compared dose distribution. ΔD_M denotes dose difference tolerance, while Δd_M denotes maximum allowed distance to agreement (graph adapted from Depuydt et al. [19]).

5.2.6 DVHs Evaluation

I found it important to compare not only prediction of dose distribution but also dosimetric parameters achieved for particular structures. Gamma evaluation do not take into account how the dose is distributed in the structures and how well estimation is done in that sense. In order to compare cumulative $D_{p,cum}$ and estimated $D_{p,est}$ dose distribution in terms of doses delivered to structures I used DVHs (2.4). Dosimetric parameters in CTV, PTV and rectum were investigated. For the plan performance the most important would be dose in CTV as CTV-PTV margin was created in order to counteract setup errors. For CTV and PTV three parameters were evaluated: D_{min}, D_{mean} and $D_{98\%}$. For rectum $V_{60\ \mathrm{Gy}}$ and $V_{65\ \mathrm{Gy}}$ were evaluated. At first a difference between $D_{p,cum}$ and $D_{p,ms}$ or $D_{p,md}$ was calculated. Afterwards a statistical comparison was done in order to check if there is a significant difference between two estimation methods in terms of doses in target volumes.

5.2.7 Statistical comparison

In order to statistically compare both methods of estimation i.e. $D_{ms}(n)$ and $D_{md}(n)$ Mann-Whitney rank test was used. For each n, for each patient and each technique, a set of $D_{ms}(n)$ and $D_{md}(n)$ was prepared. That gave 25 pairs of independent results (gamma index passing rates, dosimetric parameters from DVH evaluation). According to that assumptions for Mann-Whitney rank test were met. The null hypothesis stated that there is no difference between

two methods of cumulative dose estimation. The p-value lower than 0.05 was considered as significant, i.e. showing that there is reason to reject null hypothesis in favor of alternative one.

5.3 Results

5.3.1 Gamma Evaluation

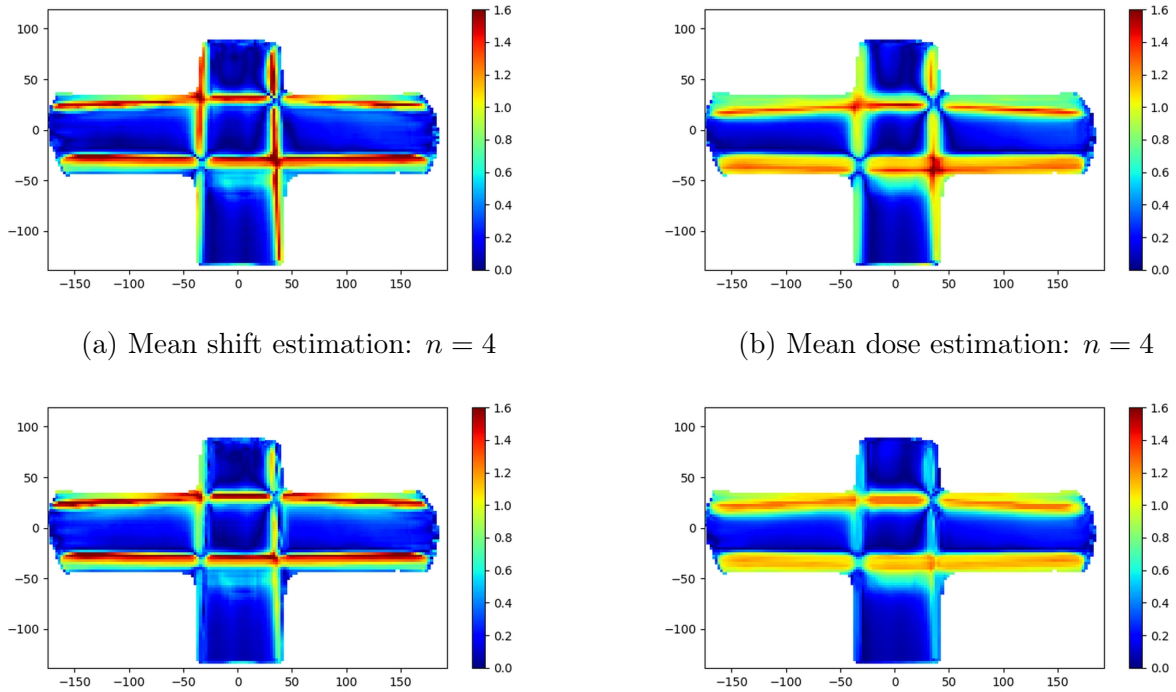
In clinical practice results of gamma evaluation are shown on 2D maps with color temperature as a scale. It has to be kept in mind that voxels with gamma bigger than 1 did not pass the criteria — i.e. the difference between dose distributions is bigger than the acceptance criteria. Similar gamma maps were prepared for performed gamma analysis of $D_{p,est}(n)$ and some examples are shown in Fig. 5.4, 5.6, 5.8. Comparisons between $n = 4$ and $n = 8$ for 3D (see Fig. 5.4), IMRT (see Fig. 5.6) show that for bigger n estimation of dose in the penumbra region (i.e. in the region close to the field border) is better. It can be explained by dose blurring increasing with n. Effect of dose blurring can be also seen for VMAT technique (see Fig. 5.8) even if it is rather impossible to define field border there. The presented comparison shows also that gamma results for $D_{p,md}(n)$ estimation have lower gamma values than $D_{p,ms}(n)$, especially in high dose gradient regions. That can be explained by a dose blurring which is incorporated in dose estimation only in $D_{p,md}(n)$ method. Evaluated volume is biggest for VMAT technique, and smallest for 3D-CRT. That is because IMRT uses more fields than 3D-CRT and VMAT uses full arcs (so there is a beam from each gantry angle). As a consequence in both techniques there is a rather big volume of normal tissue receiving low (but bigger than 10%) dose [35, 71].

5.3.1.1 Estimating with mean setup error

Results of gamma comparison for $D_{p,ms}(n)$ are shown in Fig. 5.5, standard deviation of achieved results is presented as errorbars. For 3D technique for $D_{p,ms}(n \geqslant 9)$ ($D_{p,ms}(n \geqslant 13)$) the gamma index 2 mm 2% of global (local) dose, with error, was smaller than 1 in at least 95% (90%) of analyzed voxels. For IMRT technique for $D_{p,ms}(n \geqslant 5)$ ($D_{p,ms}(n \geqslant 7)$) the gamma index 2 mm 2% of global (local) dose, with error, was smaller than 1 in at least 95% (90%) of analyzed voxels. For VMAT technique for $D_{p,ms}(n \geqslant 3)$ ($D_{p,ms}(n \geqslant 4)$) the gamma index 2 mm 2% of global (local) dose with error was smaller than 1 in at least 95% (90%) of analyzed voxels.

5.3.1.2 Estimating with mean dose distribution

Results of gamma comparison for $D_{p,md}(n)$ are shown in Fig. 5.7, standard deviation of achieved results is presented as errorbars. For 3D technique for $D_{p,md}(n \geqslant 8)$ ($D_{p,md}(n \geqslant 8)$) the gamma index 2 mm 2% of global (local) dose, with error, was smaller than 1 in at least 95% (90%) of analyzed voxels. For IMRT technique for $D_{p,md}(n \geqslant 4)$ ($D_{p,md}(n \geqslant 6)$) the gamma index 2 mm 2% of global (local) dose, with error, was smaller than 1 in at least 95% (90%) of analyzed voxels. For VMAT technique for $D_{p,ms}(n \geqslant 3)$ ($D_{p,md}(n \geqslant 4)$) the gamma index 2 mm 2% of global (local) dose with error was smaller than 1 in at least 95% (90%) of analyzed voxels.



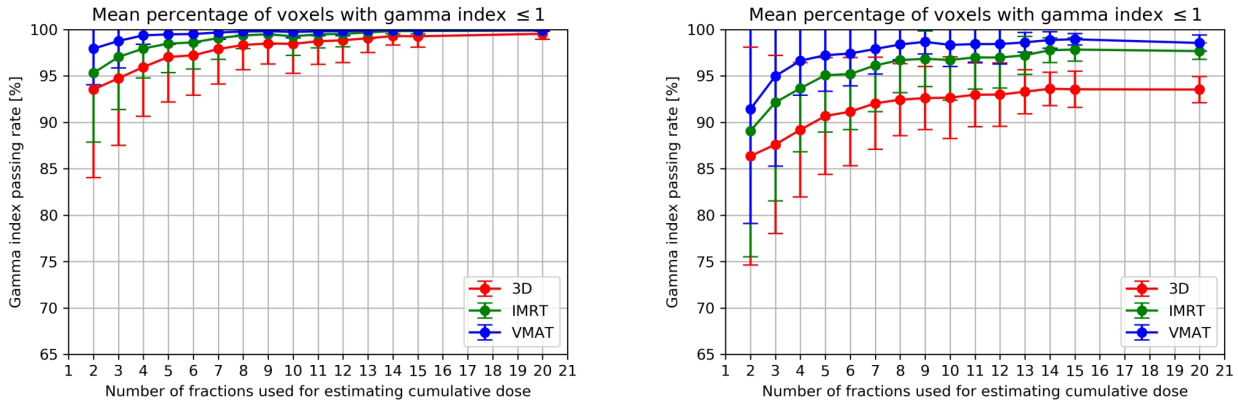
(a) Mean shift estimation: $n = 4$

(b) Mean dose estimation: $n = 4$

(c) Mean shift estimation: $n = 8$

(d) Mean dose estimation: $n = 8$

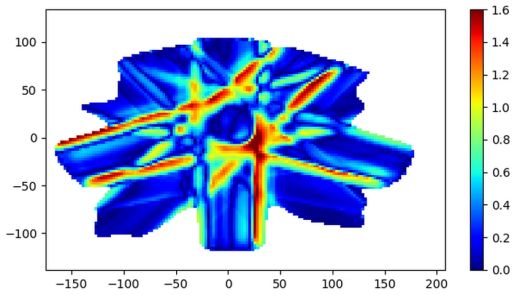
Figure 5.4: Examples of gamma evaluation result for 3D-CRT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient.



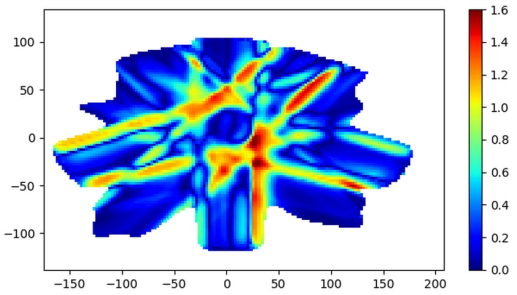
(a) global gamma

(b) local gamma

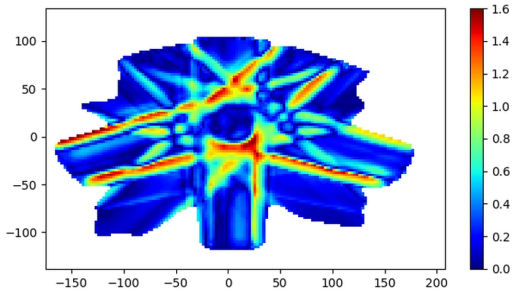
Figure 5.5: Mean percent of voxels with gamma index not greater than 1 ($\gamma \leqslant 1$) for comparison between $D_{p,cum}$ and $D_{p,ms}(n)$. Standard deviation is presented as error. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right).



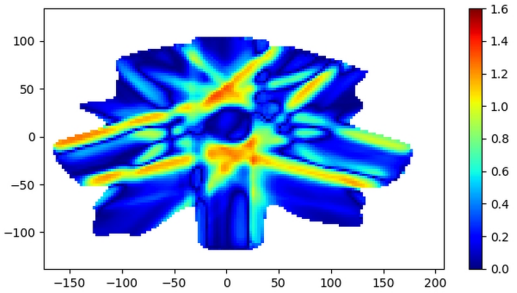
(a) Mean shift estimation: $n = 4$



(b) Mean dose estimation: $n = 4$

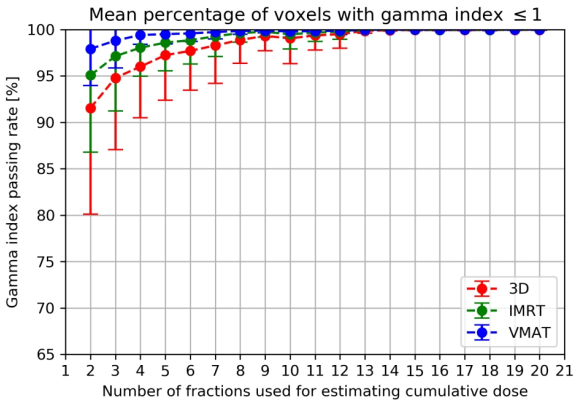


(c) Mean shift estimation: $n = 8$

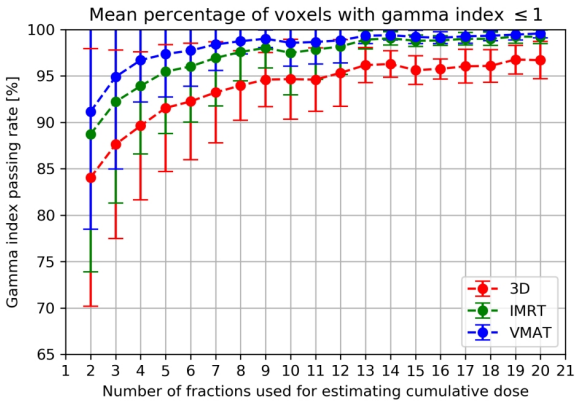


(d) Mean dose estimation: $n = 8$

Figure 5.6: Examples of gamma evaluation result for IMRT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient.

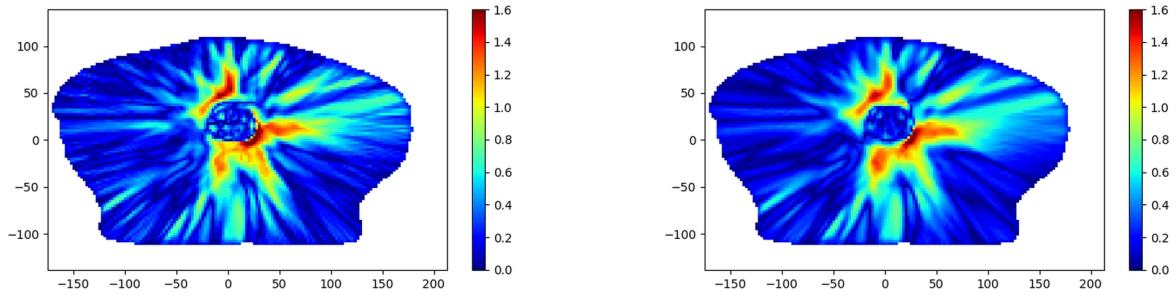


(a) global gamma



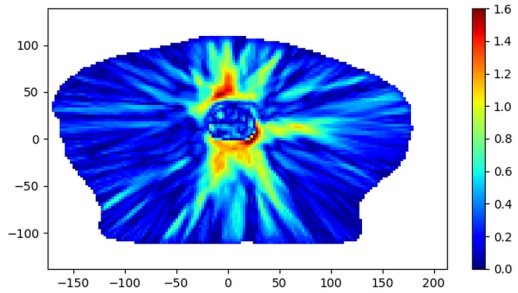
(b) local gamma

Figure 5.7: Mean percent of voxels with gamma index not greater than 1 ($\gamma \leqslant 1$) for comparison between $D_{p,cum}$ and $D_{p,md}(n)$. Standard deviation is presented as error. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right).

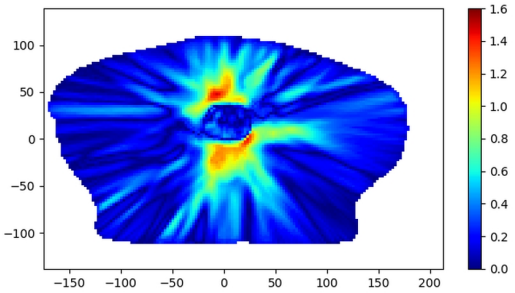


(a) Mean shift estimation: $n = 4$

(b) Mean dose estimation: $n = 4$

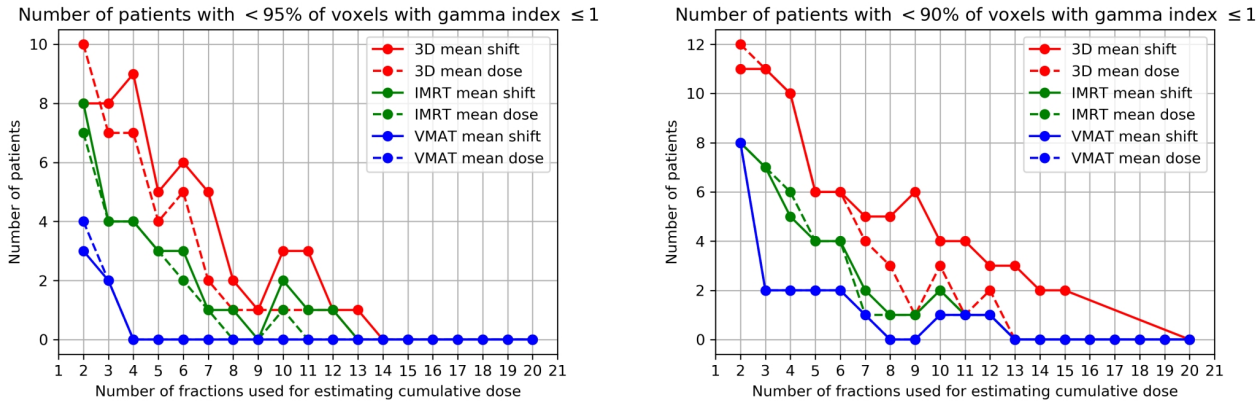


(c) Mean shift estimation: $n = 8$



(d) Mean dose estimation: $n = 8$

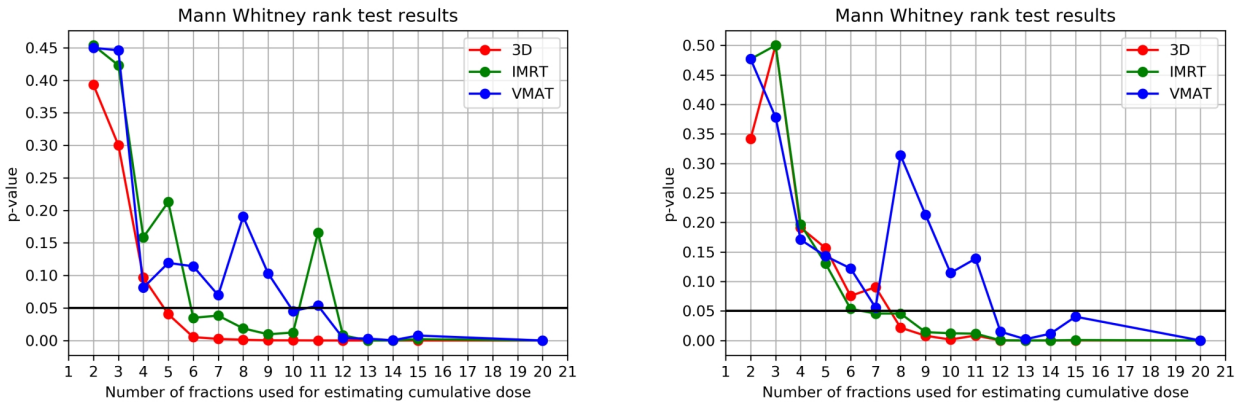
Figure 5.8: Examples of gamma evaluation result for VMAT technique. Gamma below or equal 1 (shown in blue and green) corresponds to voxels which fulfilled the dose and distance difference tolerance: $\Delta D_M = 2\%$ max dose and $\Delta d_M = 2$ mm. Difference between estimation from $n = 4$ (upper figures) and $n = 8$ (lower figures) known fractions is shown. Estimation with mean shift is shown on the left, estimation with mean dose is shown on the right side of figure. All presented gamma distributions were calculated for the same patient.



(a) global gamma

(b) local gamma

Figure 5.9: Number of patients with gamma index not greater than 1 ($\gamma \leqslant 1$) in less than 95% (90%) of voxels for global (local) gamma evaluation. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right).



(a) global gamma (b) local gamma

Figure 5.10: Results of Mann-Whitney test performed for different n fractions used in dose estimation. Gamma index passing rate for $D_{p,ms}(n)$ and $D_{p,md}(n)$ were compared. Gamma parameters were set to 2 mm 2% of global dose (left) or local dose (right).

5.3.1.3 Comparison of estimation methods

In order to compare both methods of estimating $D_{p,cum}$ number of patients who did not pass the gamma evaluation — i.e. for whom gamma index 2 mm 2% of global (local) dose was in less than 95% (90%) of voxels (see Fig. 5.9) — was also examined. It can be seen that in general, with increasing number of fractions n taken into cumulative dose estimation, the number of patients not passing gamma evaluation decreases. The effect is independent on the technique. Results are similar for IMRT and VMAT technique, although for 3D technique difference might be observed especially in local gamma evaluation.

Mann-Whitney rank test was performed in order to test the null hypothesis of no difference between two methods of dose estimation (see Fig. 5.10). For global gamma evaluation we have basis to neglect the null hypothesis for $n \geqslant 5$ for 3D-CRT technique ($p \leqslant 0.05$). Also for IMRT technique (apart from $n = 11$) we can neglect the null hypothesis for $n \geqslant 6$. For VMAT technique for $n < 10$ we have no basis to reject the null hypothesis. It seems to be connected with the bigger irradiated volume in case of VMAT technique. That technique seems to be most robust for setup errors in terms of geometrical dose distribution. Dose blurring which may be estimated with bootstrap methodology is more important in 3D-CRT where field borders are seen and steeper dose gradients are achieved in the penumbra region (see Fig. 5.11).

5.3.2 DVHs Evaluation

5.3.2.1 Comparison of doses in CTV

Fig. 5.12 presents results of a comparison between dose in CTV for cumulative dose distribution $D_{p,cum}$ and estimated dose distribution: $D_{p,ms}$ and $D_{p,md}$. It can be clearly seen that error (which was the SD) does not change or change only slightly for $D_{p,ms}$. Also mean difference is rather constant despite number of fractions n used for estimation. On the other hand SD of difference between $D_{p,md}$ and $D_{p,cum}$ is decreasing with increasing n. For D_{min} and $D_{98\%}$ also mean dose difference is decreasing with increasing n. It has to be stressed that negative value of dose difference means that $D_{p,cum}$ was bigger than D_{est}. In case of using this type of

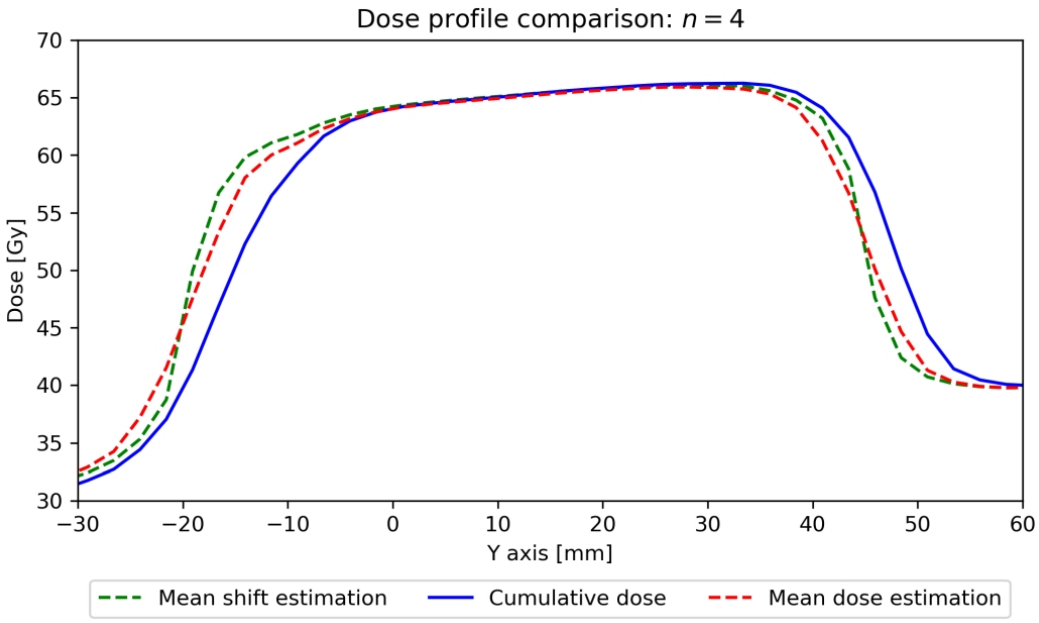


Figure 5.11: Comparison of profiles for $D_{p,cum}$, $D_{p,ms}(n)$ and $D_{p,md}(n)$. It can be seen that $D_{p,ms}(n)$ has the steepest dose gradient in penumbra region.

estimation and evaluation during routine planning it would be very important. For estimated $D_{98\%}$ lower than one for $D_{p,cum}$ but still acceptable the actually delivered dose would be higher than estimated one. Smallest dose differences for D_{min} and $D_{98\%}$ are seen for 3D technique followed by IMRT and VMAT. For D_{mean} the smallest dose difference (and positive) is seen for IMRT, 3D and VMAT are similar.

5.3.2.2 Comparison of doses in PTV

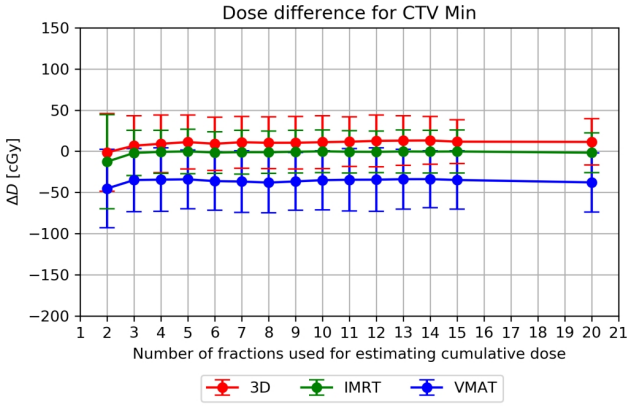
Fig. 5.13 presents results of comparison between cumulative dose distribution $D_{p,cum}$ and estimated dose distribution: $D_{p,ms}$ and $D_{p,md}$ in terms of doses in PTV. Contrary to Fig. 5.12 it can be seen that mean dose difference for PTV is positive thus suggesting that $D_{p,cum}$ was lower than D_{est}. That should be kept in mind in case of clinical practice. The observations about almost constant mean and SD of dose difference for $D_{p,ms}$ despite n remains the same. For $D_{p,md}$ the mean dose difference as well as SD converges to zero with increasing n.

5.3.2.3 Comparison of doses in rectum

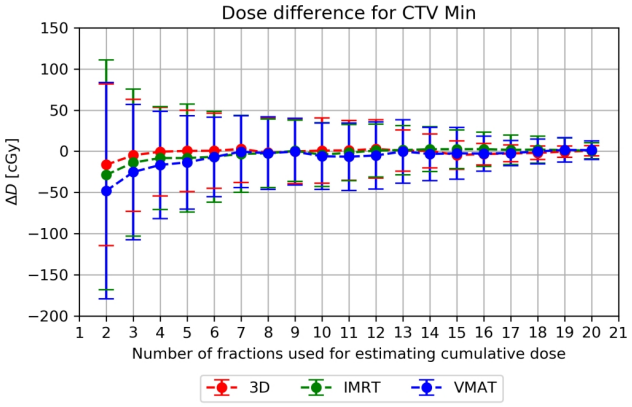
Comparison between cumulative dose distribution $D_{p,cum}$ and estimated dose distribution: $D_{p,ms}$ and $D_{p,md}$ in terms of doses in rectum is presented in Fig. 5.14. Mean and SD of volume differences for both $V_{60\text{ Gy}}$ and $V_{65\text{ Gy}}$ decrease with increasing number of fractions used in estimation for $D_{p,md}$.

5.3.2.4 Comparison of estimation methods

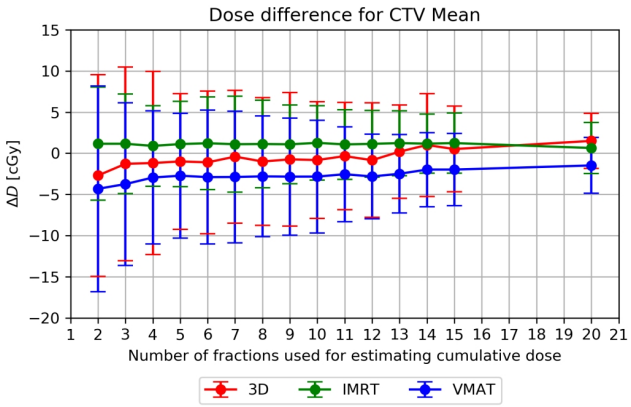
Results of statistical comparison between two methods of estimation can be seen in Fig. 5.15-5.16. Null hyphotesis of no differences between the estimation methods cannot be rejected



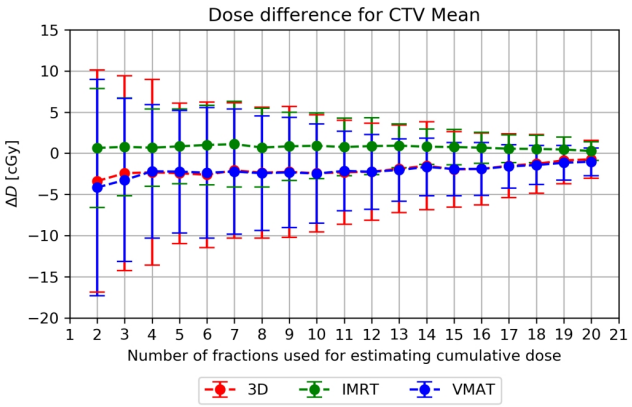
(a) Mean shift estimation: D_{min}



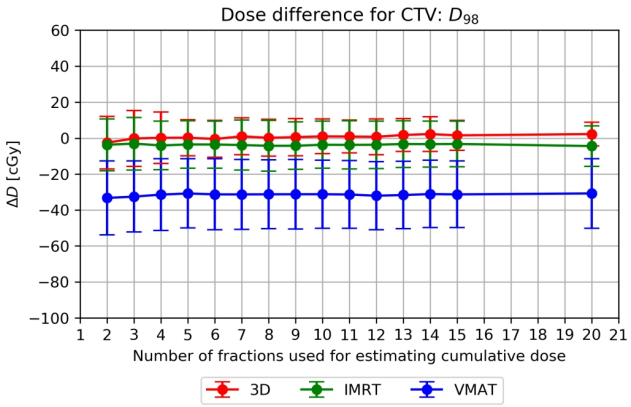
(b) Mean dose estimation: D_{min}



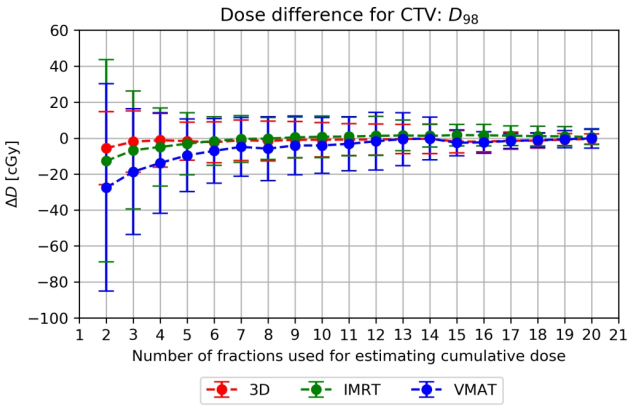
(c) Mean shift estimation: D_{mean}



(d) Mean dose estimation: D_{mean}

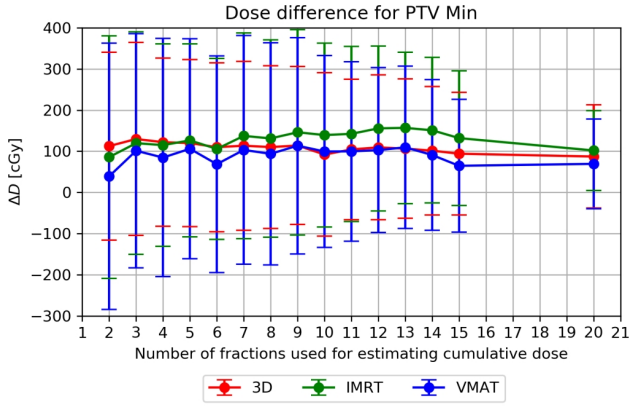


(e) Mean shift estimation: $D_{98\%}$

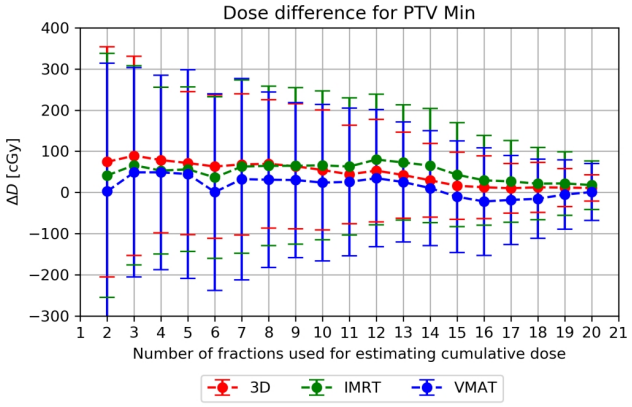


(f) Mean dose estimation: $D_{98\%}$

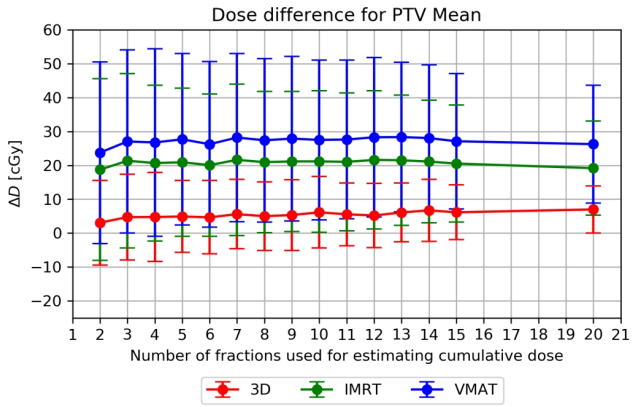
Figure 5.12: Comparison of dosimetric parameters for CTV between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in D_{min}, D_{mean} and $D_{98\%}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}.



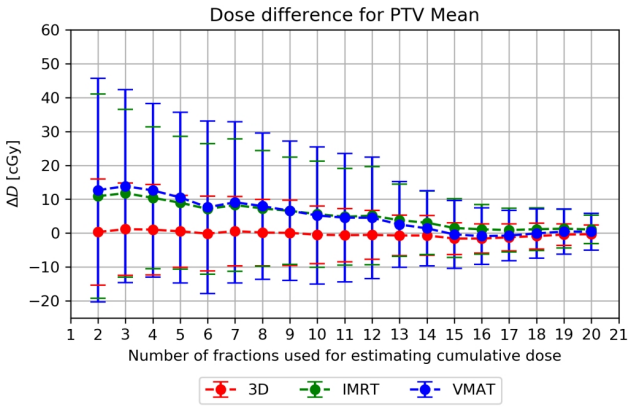
(a) Mean shift estimation: D_{min}



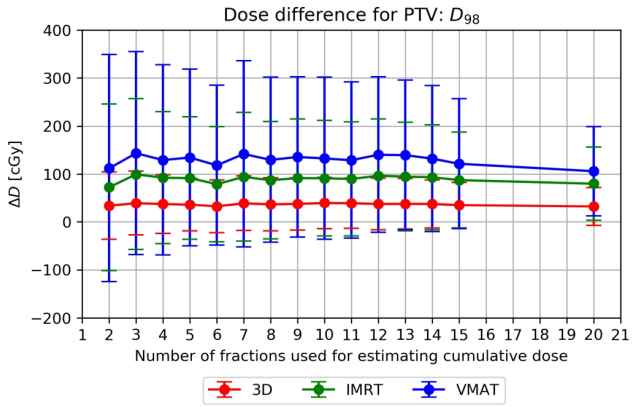
(b) Mean dose estimation: D_{min}



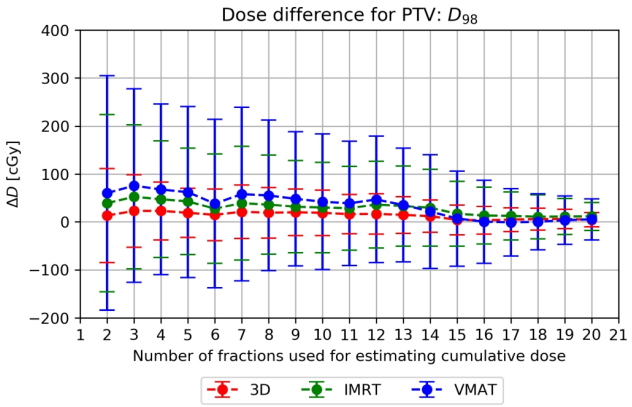
(c) Mean shift estimation: D_{mean}



(d) Mean dose estimation: D_{mean}

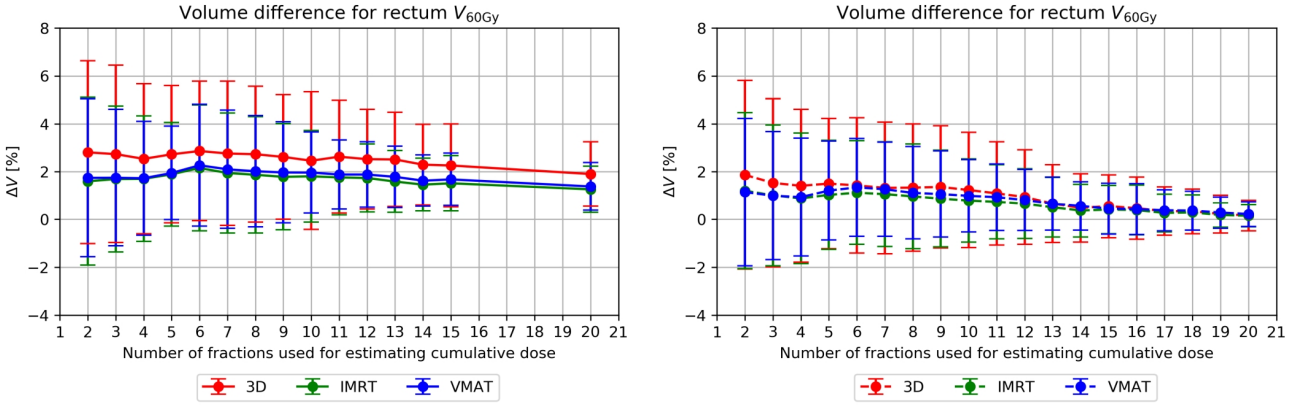


(e) Mean shift estimation: $D_{98\%}$



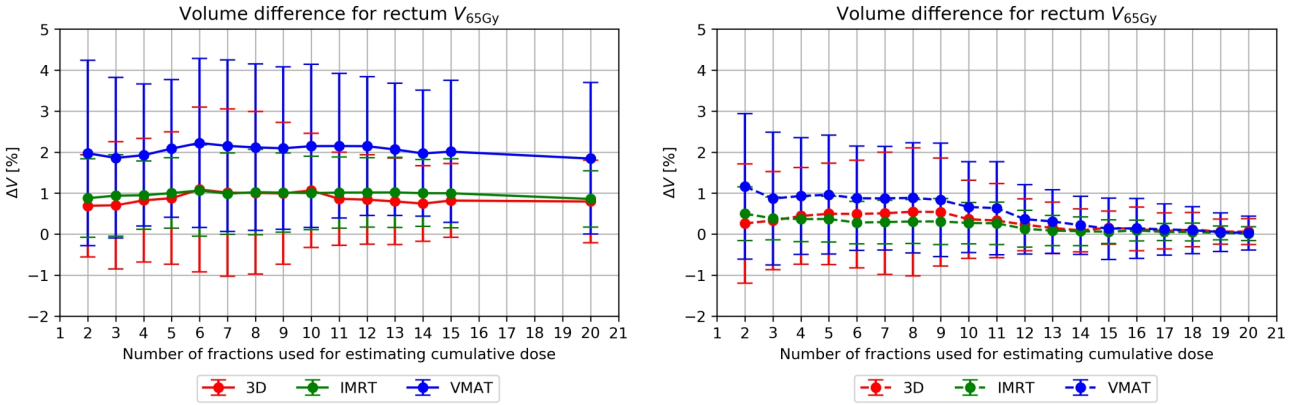
(f) Mean dose estimation: $D_{98\%}$

Figure 5.13: Comparison of dosimetric parameters for PTV between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in D_{min}, D_{mean} and $D_{98\%}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}.



(a) Mean shift estimation: $V_{60\ \mathrm{Gy}}$

(b) Mean dose estimation: $V_{60\ \mathrm{Gy}}$

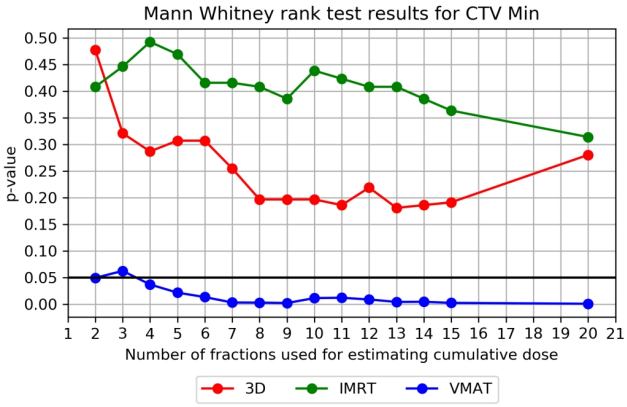


(c) Mean shift estimation: $V_{65\ \mathrm{Gy}}$

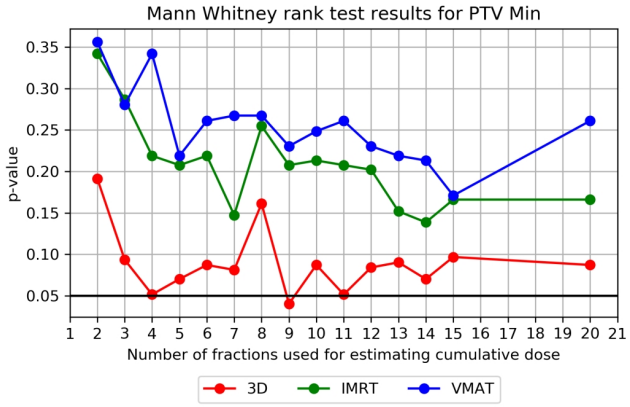
(d) Mean dose estimation: $V_{65\ \mathrm{Gy}}$

Figure 5.14: Comparison of dosimetric parameters for rectum between $D_{p,cum}$ and $D_{p,ms}$ (left) or $D_{p,md}$ (right). Differences in $V_{60\ \mathrm{Gy}}$ and $V_{65\ \mathrm{Gy}}$ are shown. Number of fractions used for dose estimation are shown in x axis. Negative values mean that $D_{p,cum}$ was bigger than D_{est}.

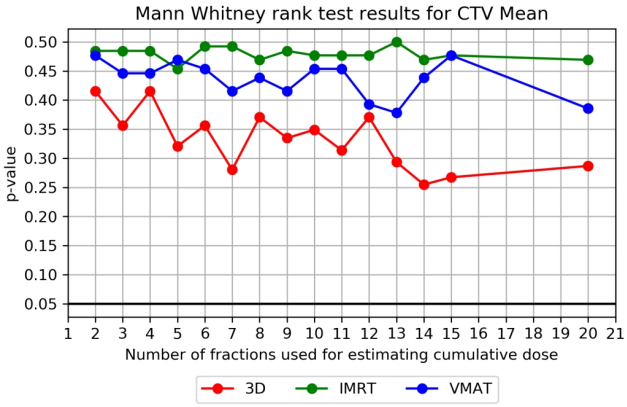
for CTV D_{mean} for all techniques. The null hypothesis has to be rejected for CTV D_{min} and $D_{98\%}$ for VMAT technique for n bigger than 4 and 2 respectively. Combining that result with Fig. 5.12 shows that the cumulative dose distribution is better estimated with $D_{p,md}$. Estimation is significantly different for IMRT and VMAT technique for PTV D_{mean} and $n \geqslant 3$ as well as for rectum $V_{65\ \mathrm{Gy}}$ and $n \geqslant 11$.



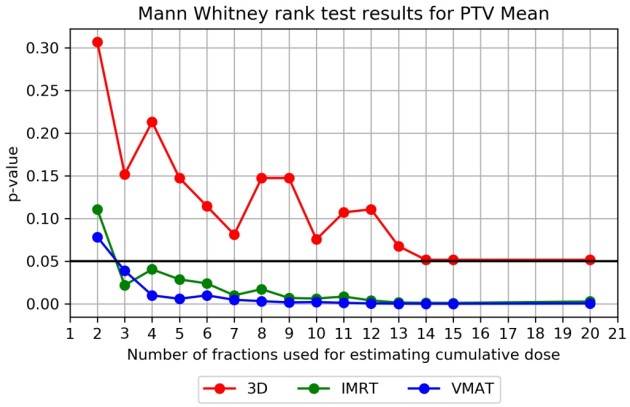
(a) CTV: D_{min}



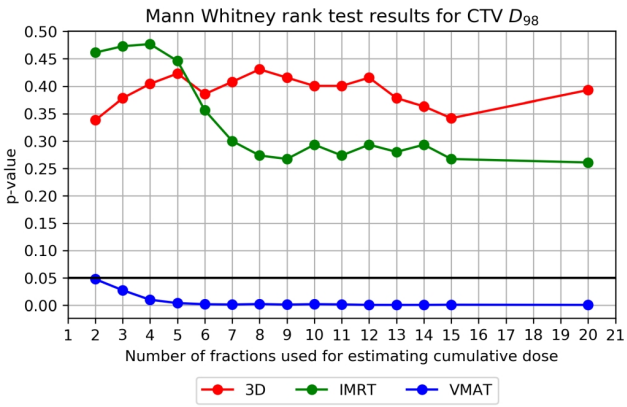
(b) PTV: D_{min}



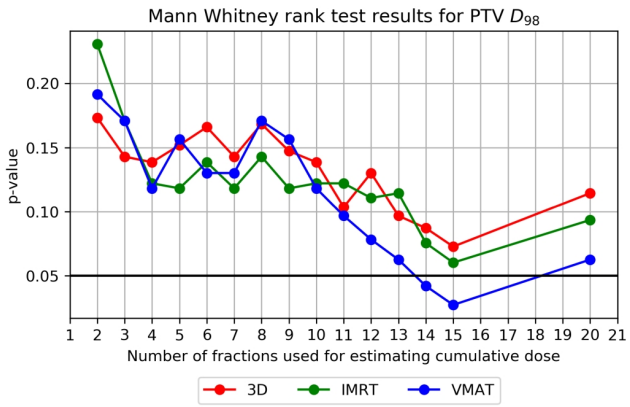
(c) CTV: D_{mean}



(d) PTV: D_{mean}

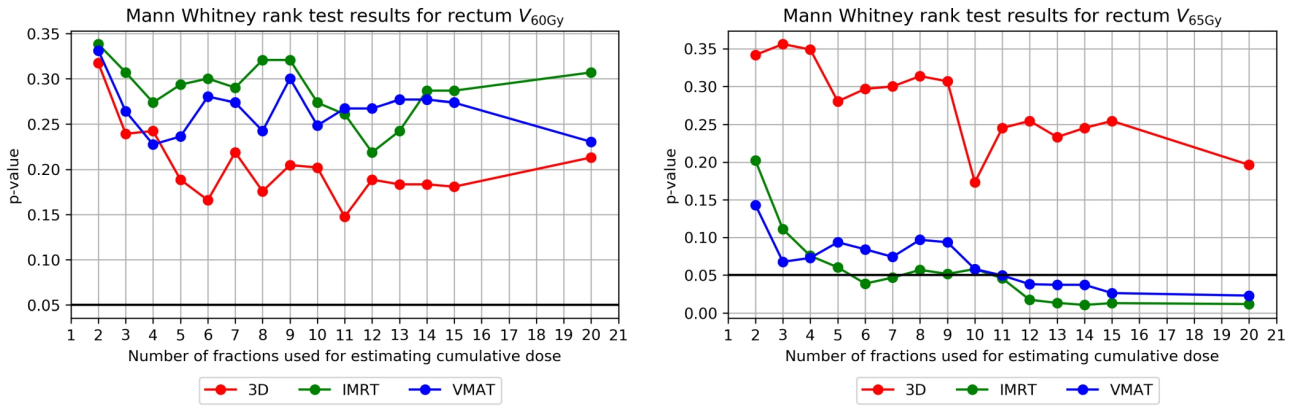


(e) Mean shift estimation: $D_{98\%}$



(f) Mean dose estimation: $D_{98\%}$

Figure 5.15: Mann-Whitney sign rank test results for dose parameters in CTV and PTV.



(a) rectum: $V_{60\ \mathrm{Gy}}$ (b) rectum: $V_{60\ \mathrm{Gy}}$

Figure 5.16: Mann-Whitney sign rank test results for dose parameters in rectum.

5.4 Discussion

The effect of setup errors and target/organ motion on dose distribution was investigated by many groups and summarized in [6].

Convolution method of incorporating setup uncertainties into the plan evaluation has been widely used and discussed [11, 12, 89, 90]. That method stands also behind the van Herk's recipe for margin calculation (see Section 2.2.1). Craig et al. presented limitations of these methodology [11, 12]. One limitation is that the convolution is usually done on the basis of static dose cloud and so it does not incorporate anatomical changes which would influence the dose distribution. Secondly, the convolution method assumes infinite number of infinitesimally small fractions which is not the case in practice. The solution presented in this part of my work deals with both these problems. The dose distribution was always calculated with the isocenter shift and therefore the dose per fraction was not a fixed dose cloud. The limitation of this study was the fixed CT image used. It is well known that for prostate irradiation changes in target and OARs positions and shapes are observed. Still I found it important to investigate with solid anatomy whether proposed methodology might be used. For further investigations daily CBCT could be used in order to incorporate not only differences in dose distributions due to beam shift but also due to anatomical changes.

Nowadays much effort is put into designing margin-less planning and patient specific treatment. First approach is mainly based on probability distributions of possible setup errors or organ motions [5, 27]. In second approach known target positions and shapes are incorporated to create a patient specific margin [95] or to create plans for different patient geometries in plan of the day approach [31, 41].

The proposed method of estimating the cumulative dose distribution may be used in order to create a patient specific margin-less plan. Moreover, that methodology may be used for the plan evaluation in daily practice without the need of new treatment planning software.

Chapter 6

Proposal of a new off-line verification protocol

6.1 Background

No Action Level (NAL) Protocol [14], which was described in Section 2.3 is an example of an off-line verification protocol (OVP). It provides a reduction in systematic error while keeping the workload on a reasonably low level. In NAL one has to measure setup errors in n first fractions (usually $n = 3$ or $n = 4$ [7, 14]) and then calculate mean value of setup error for each axis which would be used as a correction value for all upcoming fractions. NAL doesn't take into account the possible time and trend therefore an eNAL was introduced [16]. As Bortfeld [7] noticed the NAL does not always end up with a good estimation of systematic setup error. That is why it can happen that pure (i.e. not corrected) setup errors would have smaller setup error than those corrected with NAL (see Fig. 6.1).

The quality of NAL performance depends on the systematic setup error estimation. As de Boer mentioned in [14]: "final distribution of systematic errors is basically independent of the initial distribution of systematic errors, and determined only by SD of the random errors and the number of measurements n per patient from which the systematic error of patient is estimated." It is known from statistics that for larger SD the estimate of mean value is usually worse with the same number of samples used. As a consequence it can be expected that estimate of systematic setup errors for patients with large random error variation (so called large movers) can have bigger uncertainty. From clinical experience it is also known that some patients require an on-line verification for whole treatment. That is why I decided to check the reliability (the accuracy) of mean setup error estimation calculated after first few fractions of treatment.

The aim of this part of my work was to propose a new OVP taking into account the uncertainty of systematic setup error estimation. In other words the idea was to identify problematic patients from the whole cohort. I define a problematic patient here as the one with high probability of bad estimation of mean setup error in the first fractions of treatment. These patients would most probably have larger setup errors variations and that is why they are called large movers.

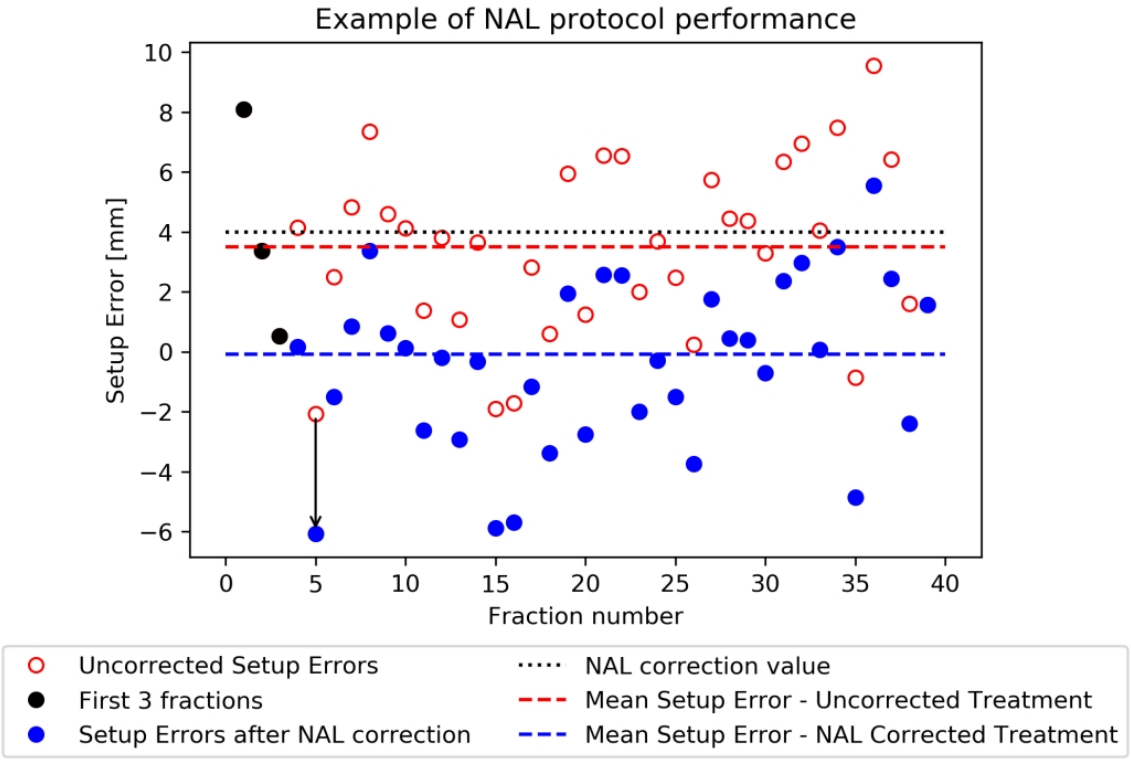


Figure 6.1: Example of of wrong estimation of mean setup error in NAL protocol. Arrow shows an example shift between corrected an uncorrected setup error.

6.2 Material and methods

6.2.1 Individualized No Action Level Protocol (iNAL)

The proposed individualized No Action Level (iNAL) protocol starts as NAL. Images are taken in the first 3 fractions, without any correction being applied. A setup correction factor is calculated after the first 3 fractions. The difference in iNAL protocol is that starting from the 4th fraction the estimation of standard error of the mean from all previous fractions is also examined. That value is a measure of uncertainty of the mean value. Due to the fact that standard deviation is calculated from the sample of population the unbiased estimator of the standard deviation is used:

$$SEM_{\overline{m_p}(n)} = \frac{SD_p}{\sqrt{n}} = \sqrt{\sum_{f=1}^{n} \frac{(m_{pf} - \overline{m_p})^2}{n(n-1)}} \tag{6.1}$$

where:

$SEM \rightarrow$ standard error of mean

$n \rightarrow$ number of fractions taken into calculation (i.e. the number of fractions with images)

The $SEM_{\overline{m_p}}(n)$ is compared with some arbitrarily chosen value called flag. If it is lower than the flag then NAL protocol is applied until the end of the treatment. If it is not, the correction is applied only in upcoming fraction with image being taken. Afterwards the procedure is done

again and again — until $SEM_{\overline{m_p}}(n)$ is lower than the flag. After each new image taken a correction as well as $SEM_{\overline{m_p}}(n)$ are recalculated.

6.2.2 Cumulative Average Protocol (CA)

In NAL protocol n fractions would be irradiated without setup error correction and all other fractions will have it applied. In iNAL protocol all patients have correction applied after the 3rd fraction, although for some patients the correction value would be recalculated. Therefore the question arises if the difference in protocols performance might be attributed to the number of uncorrected fractions. That is why in order to provide a more fair comparison between protocols I also tested the Cumulative Average (CA) OVP protocol.

The CA protocol started in the same manner as NAL. Three first fractions were done without any corrections and the mean value of setup error was calculated. Afterwards the correction was applied but also the image was taken until fraction m. After each fraction the correction value was recalculated as an average from all previous fractions. After fraction m correction was than applied to all upcoming fractions.

6.2.3 Synthetic Data (Monte Carlo patients)

In order to test different possible groups of patients I used different combinations of $\Sigma \in \{1, 2, 3\}$ mm, $\sigma \in \{1, 2, 3, 4\}$ mm and $SD_{SD} \in \{0.5, 1.0, 1.5, 2.0\}$ mm. For some simulations SD_{SD} was set to 0.26σ [91]. For each combination of Σ, σ and SD_{SD} 10^4 patients were simulated. I have tested the 25 and 39 fractions treatments.

Due to the fact that the aim of proposed protocol was to detect patients with large random error deviation I was very interested in testing the iNAL protocol for such patients. In order to look closely on large movers I also simulated them. In that case I set the condition that SD_p has be in upper 5% percentile of $G(\sigma, SD_{SD})$ distribution. Having the same number of patients for whole population and large movers I could provide a fair comparison between them.

6.2.4 Chi-square distribution theoretical background

In statistics the χ^2 test is used to check the accuracy of standard deviation calculation from some samples instead of whole population. The statistic ${\chi_{STAT}}^2$ is used in that case (see Eq. 6.2).

$$ {\chi_{STAT}}^2 = \frac{(k-1) \cdot S^2}{{\sigma_o}^2} \tag{6.2} $$

where:

$k \rightarrow$ sample size

$S^2 \rightarrow$ sample variance

${\sigma_o}^2 \rightarrow$ population variance as stated in the null hypothesis

In iNAL protocol the aim is to detect patients with large random error variation. It might happen that having relatively small population variance ${\sigma_o}^2$ the sample variance calculated from first few fractions (i.e. from small sample size) would be big. For that patients, in iNAL protocol, an additional imaging would be done. Of course the decision of the additional imaging would depend on the value of arbitrary chosen flag used in the protocol. On the basis of χ^2

distribution it is possible to predict the number of images which will be done in the iNAL protocol depending on the flag.

In case of the iNAL the Eq. (6.2) has to be rewritten due to the fact that we use standard error of mean ($SEM_{\overline{m_p}}(n)$) instead of variance to compare with the flag. In other words we are interested in probability of finding the $SEM_{\overline{m_p}}(n)$ for n samples equal to the flag in case that the population standard error of mean is SEM_o.

$$\chi_{STAT}{}^2(SEM_o(n_f)) = \frac{(n_f-1)\cdot\dfrac{S^2}{n_f}}{\dfrac{{\sigma_o}^2}{n_f}} = \frac{(n_f-1)\cdot \text{flag}^2}{{SEM_o}^2} \tag{6.3}$$

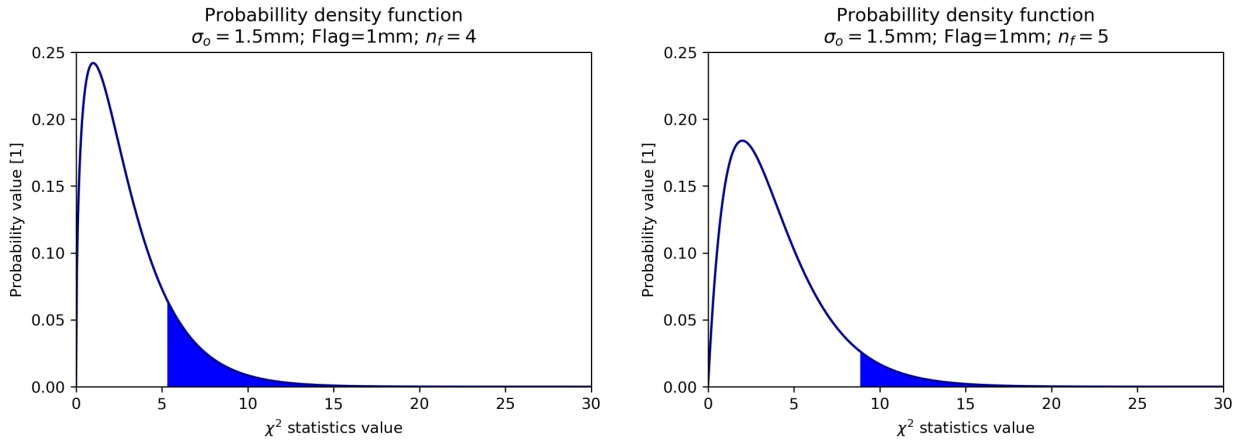
In Eq. (6.3) n_f is the number of fractions with images for a particular patient (number of samples in statistical terminology) and ${\sigma_o}^2$ is the true variance of setup errors for examined population of patients. In case of Monte Carlo patients it means that setup errors were randomly chosen from the normal distribution with ${\sigma_o}^2$ used as variance.

In iNAL protocol the number of images for a particular patient depends on the comparison between $SEM_{\overline{m_p}}(n)$ and the flag. That is why in order to calculate the probability that patient would have n_f fractions with images we have to calculate the area under the χ^2 distribution. The example of shapes of Chi-square distributions for different number of degrees of freedom (different n_f) are shown in Fig. 6.2. Due to the fact that our interest is in calculating the probability of having exactly n_f fractions with images we have to subtract the probability calculated for (n_f-1) fractions from the one calculated for n_f. This cannot be done strictly because of the change in shape of Probability Density Function (PDF) as it is shown in Fig. 6.2. The desired estimation can be done by calculating the Cumulative Distribution Function (CDF)[1] of χ^2_{STAT} for $\dfrac{{\sigma_o}^2}{n_f-1}$ and $\dfrac{{\sigma_o}^2}{n_f}$ (see Fig. 6.3). The white area in Fig. 6.3 is equal to $1-\text{CDF}(\chi_{STAT}{}^2(SEM_o(n_f)), n_f-1)$. That area corresponds to the probability that patient had less than n_f fractions with images. Finally it can be written:

$$\begin{aligned}\Delta_{n_f} = &\left[\text{CDF}\left(\frac{(n_f-1)\cdot\text{flag}^2}{\dfrac{{\sigma_o}^2}{n_f}}, n_f-1\right) - \text{CDF}\left(\frac{(n_f-1)\cdot\text{flag}^2}{\dfrac{{\sigma_o}^2}{n_f-1}}, n_f-1\right)\right] \\ &\times\frac{1-\sum_{i=3}^{n_f-1}\Delta_{n_f}}{1-\text{CDF}\left(\dfrac{(n_f-1)\cdot\text{flag}^2}{\dfrac{{\sigma_o}^2}{n_f-1}}, n_f-1\right)}\end{aligned} \tag{6.4}$$

The χ^2 methodology described above may be used in order to determine the average number of images for the specified flag used. In that way it would be possible to predict the workload needed to apply the iNAL protocol.

[1] notation $\text{CDF}(\chi_{STAT}{}^2, dof)$ is used to determine the Cumulative Distribution Function for $\chi_{STAT}{}^2$ and dof degrees of freedom



(a) (b)

Figure 6.2: Probability density functions of χ^2 distribution with different degrees of freedom: 3 ($n_f = 4$) — left and 4 ($n_f = 5$) — right. The shaded area shows the probability of observing $SEM_{\bar{m}_p}$ equal or bigger than the flag.

In order to investigate relations between standard deviation of setup error and required number of fractions with images a MC simulation of 10^4 patients setup errors was done. In that case SD_p was taken exactly from the set of $\{0.5, 1.0, ..., 9.5, 10.0\}$ mm. In other words for each patient SD_p was kept the same. Systematic error was taken from the normal distribution $G(0, \Sigma)$ with $\Sigma \in \{1, 2, 3\}$ mm.

6.2.5 Receiver Operating Characteristic curve

Receiver Operating Characteristic (ROC) curve is a well known characteristic used mostly in diagnostic [88]. The aim of this curve is to check the sensitivity and specificity of the method. Assume that we would like to detect large movers. In population of patients some of them would be large movers and some not. We would like to find some classifier which would be able to detect patients that really are large movers — that would be True Positive (TP) group. It might happen that we would detect patients who were not large movers — that would be False Positive (FP) group. On the other hand all patients that are not large movers and were not detected would be True Negative (TN) group. Last, the False Negative (FN) group would consist of those patients who were large movers but were not detected. The sensitivity is a fraction of positive occurrences (here: large movers) correctly classified called the True Positive Rate (TPR). The specificity is defined as a fraction of negative occurrences correctly classified so it is True Negative Rate (TNR):

$$\text{sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6.5}$$

$$\text{specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6.6}$$

The ROC curve is a plot of TPR, i.e. sensitivity, against False Positive Rate (FPR) [36]. It

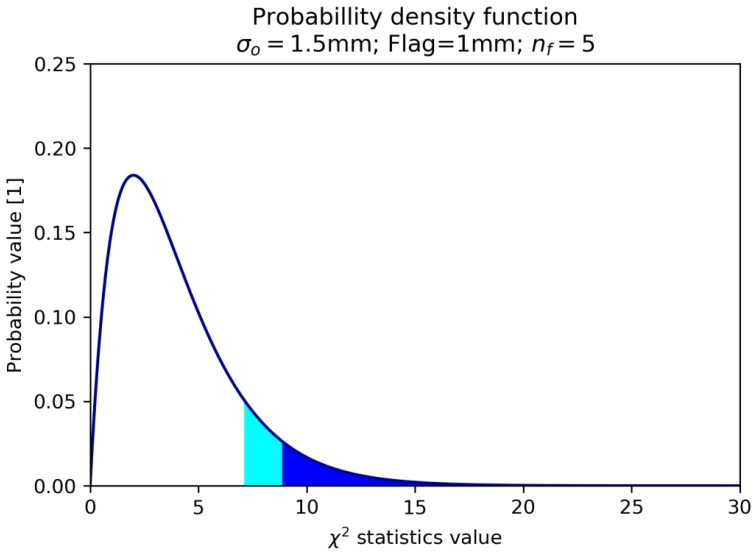


Figure 6.3: Probability density function of χ^2 distribution with 4 degrees of freedom ($n_f = 5$). The whole shaded area shows the probability of observing $SEM_{\overline{m}_p}$ equal or bigger than the flag if the $SEM_o(n_f) = \frac{{\sigma_o}^2}{(n_f - 1)}$. The blue area shows the probability of observing $SEM_{\overline{m}_p}$ equal or bigger than the flag if the $SEM_o(n_f) = \frac{{\sigma_o}^2}{n_f}$. Cyan area shows the probability of having exactly n_f fractions with images.

should be emphasized that:

$$\text{FPR} = 1 - \text{specificity} = 1 - \text{TNR} = 1 - \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{\text{FP}}{\text{TN} + \text{FP}} \tag{6.7}$$

Due to the fact that we want to detect correctly large movers we would like to maximize TPR (sensitivity) value. In the same time we would like to maximize TNR (specificity) in order to correctly not detect patients who are not large movers. It is the same with minimizing the FPR. That is why closer the ROC curve is to the top left corner better the detection method is.

Area under ROC curve (AUC) is used to check the test method accuracy. It gives information on probability of correctly classified positive occurrences [36]. The intuitive guide for the tested method classifying accuracy can be as follows:

- AUC between 0.90 and 1 means excellent test accuracy
- AUC between 0.80 and 0.90 means good test accuracy
- AUC between 0.70 and 0.80 means fair test accuracy
- AUC between 0.70 and 0.60 means poor test accuracy
- AUC below 0.60 means that test fails

The classifier used in this work was the standard error of the mean setup error $SEM_{\overline{m_p}}$ (see Eq. 6.1). The large movers were defined as those patients for whom the unbiased standard deviation SD_p (see Eq. 2.4) was at least equal to the value of 90th percentile of $G(\sigma, SD_{SD})$.

Analysis of ROC curves was done for whole Monte Carlo patients groups (including all possible SD_p).

6.2.6 Evaluation and comparison of protocols

Comparison of proposed iNAL protocol with NAL and eNAL protocols was done on the basis of residual Σ (Σ_{res}). Σ_{res} is calculated in the same manner as normal Σ (see Eq. 2.6). The only difference is that setup errors after an OVP was applied are used instead of pure ones. Value of Σ_{res} influences the calculated margin size. That is why also the margin size and percentage of patients were compared.

6.3 Results

6.3.1 Estimation of average number of images in iNAL protocol

Distribution of the fraction number with imaging for patients population with same SD_p was tested for a 25 and a 39 fractions treatment. A MC simulation of the protocol was compared with theoretical values achieved with the χ^2 statistics. Different SD_p and the flag values were evaluated. Results are presented in Fig. 6.4-6.5. It can be seen that for a bigger flag value more patients have smaller number of images. With an increase in SD_p the number of patients with bigger number of fractions with images also increases. Theoretical curve fits well to the simulated data. The peak observed in the last fraction for some SD_p and flags may be explained by the need of making images in all fractions. In a treatment with more fractions such peak would disappear.

Knowing the number of patients with different number of images (Fig. 6.4-6.5) makes possible to predict the average number of images for the specified flag value. A comparison between MC simulation and theoretical prediction was done. Results are presented in Fig. 6.6-6.7. It can be seen that χ^2 statistics methodology predicts very well the average number of images for different SD_p.

Fig. 6.8-6.9 show dependence between average number of images per patients, σ and flag. With flag decreasing the average number of images increases. With increasing σ the average number of images also increases. Results do not depend on Σ.

6.3.2 Accuracy test (ROC curves analysis)

A typical shape of ROC curve is shown in Fig. 6.10a. The classification based on $SEM_{\overline{m_p}}$ works better with n_f increase (see Fig. 6.10b). That result seems to be reasonable: with increasing n_f the estimate of standard deviation and therefore also estimate of standard mean error are closer to the real value. The AUC depends mostly on σ and SD_{SD} (see Fig. 6.11). Results for different Σ and different total number of fractions F are very similar (see Tables 6.1-6.6). The classification based on $SEM_{\overline{m_p}}(n_f = 3)$ did not fail for any of tested parameters. Effect of classification was worse with decrease of SD_{SD} and increase of σ.

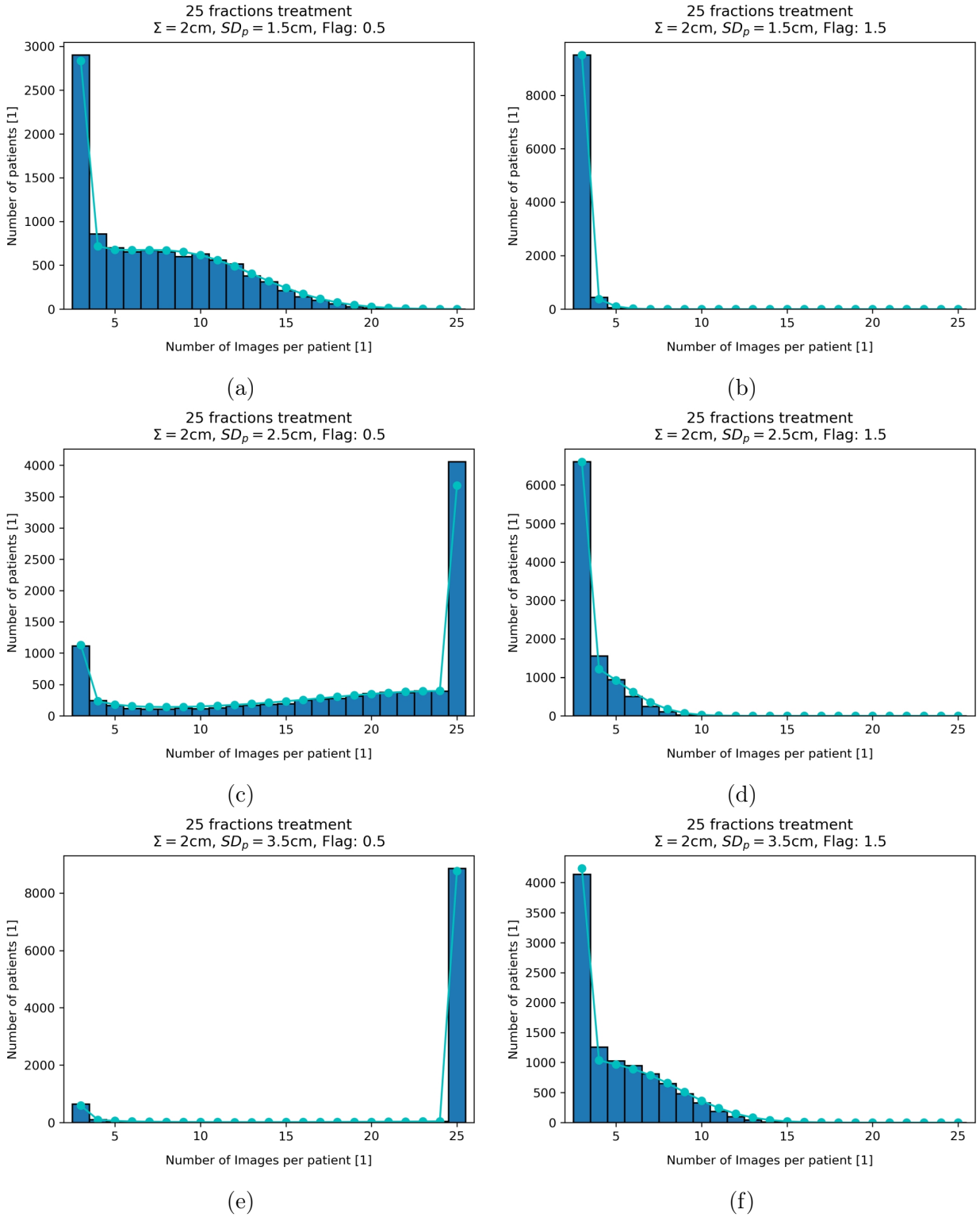


Figure 6.4: Histograms of number of patients with specified number of images. Theoretical prediction of χ^2 model is shown with cyan circles. Calculations done for 25 fractions treatment.

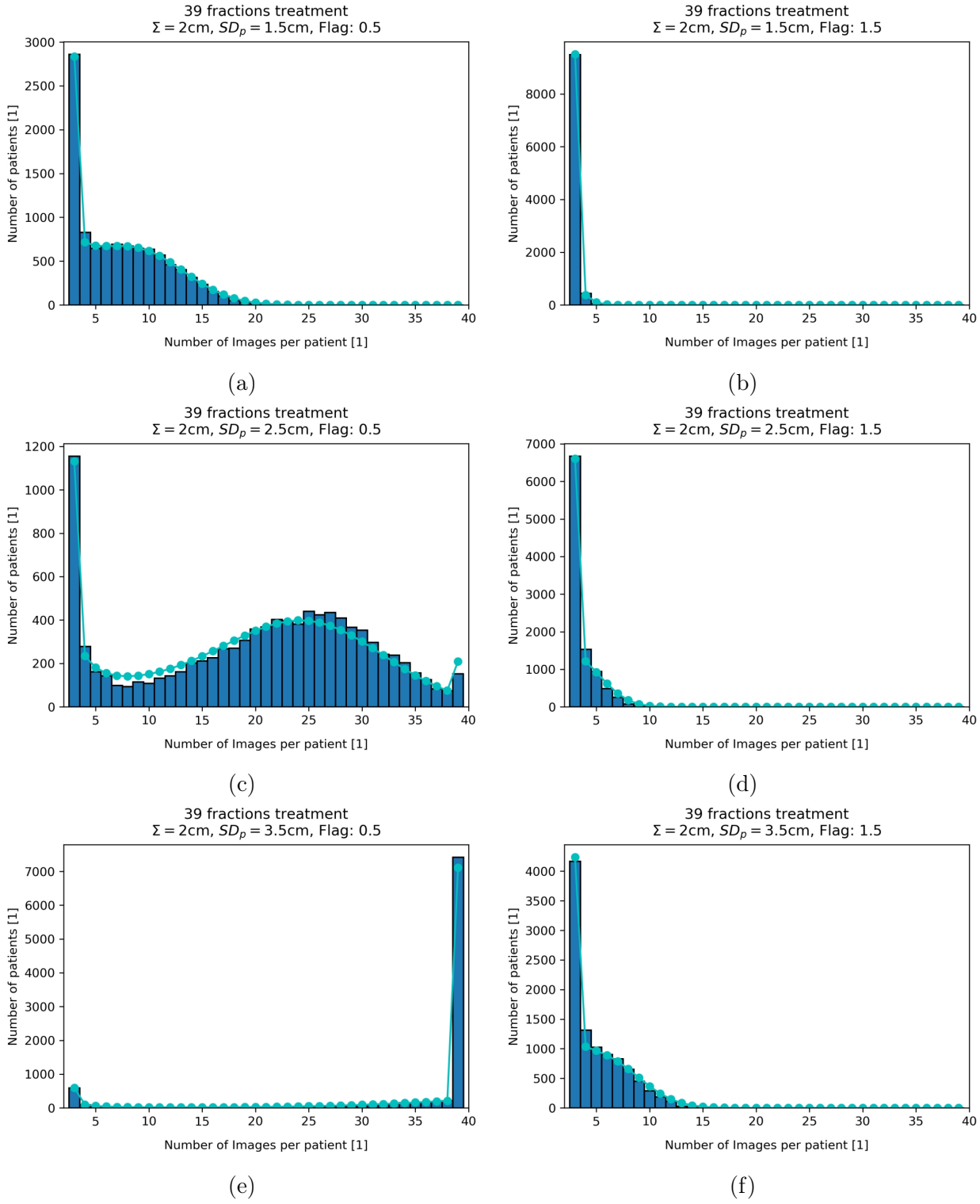


Figure 6.5: Histograms of number of patients with specified number of images. Theoretical prediction of χ^2 model is shown with cyan circles. Calculations done for 39 fractions treatment.

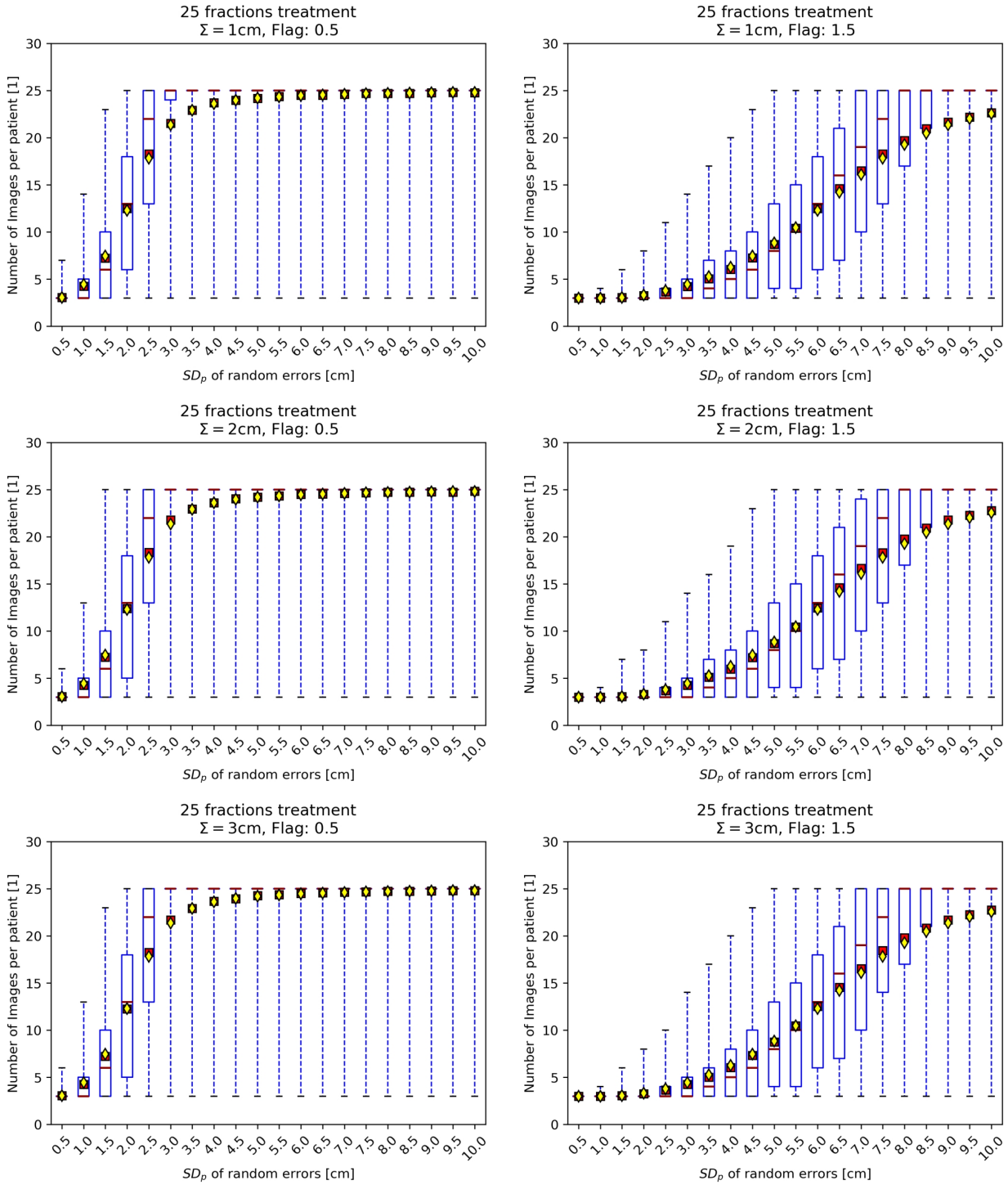


Figure 6.6: Distributions of number of images for population of patients with constant SD_p. Theoretical prediction of average number of images is shown with yellow diamond. Achieved mean value is shown with red square. Calculations done for 25 fractions treatment.

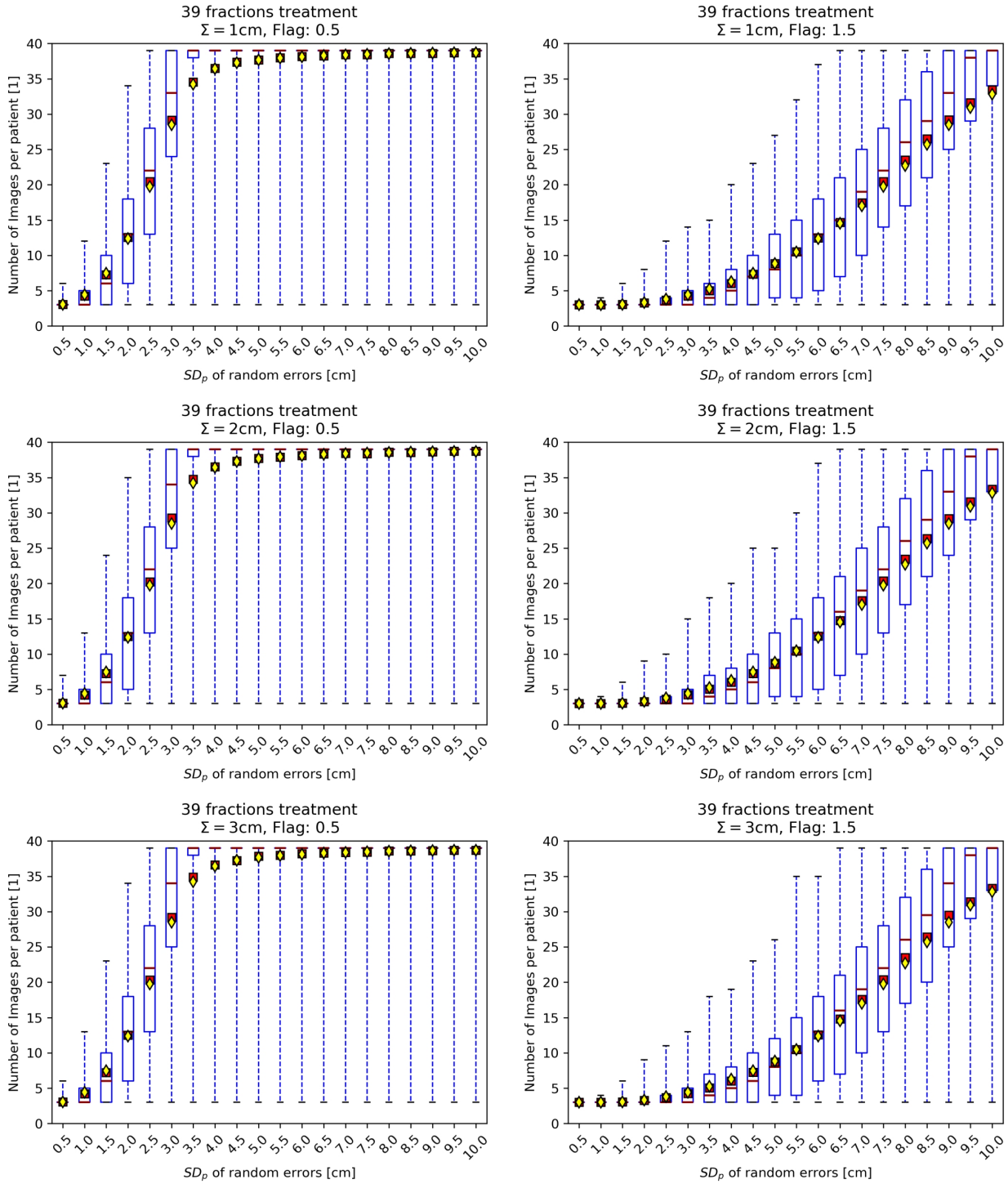


Figure 6.7: Distributions of number of images for population of patients with constant SD_p. Theoretical prediction of average number of images is shown with yellow diamond. Achieved mean value is shown with red square. Calculations done for 39 fractions treatment.

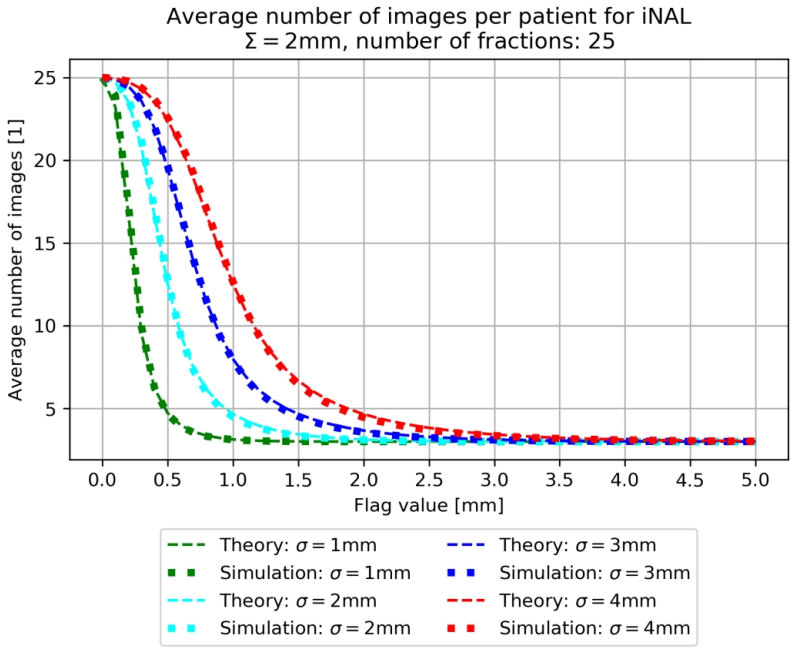


Figure 6.8: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Width of Gaussian distribution of SD_p was taken as $SD_{SD} = 0.26 \cdot \sigma$. Calculations done for 25 fractions treatment.

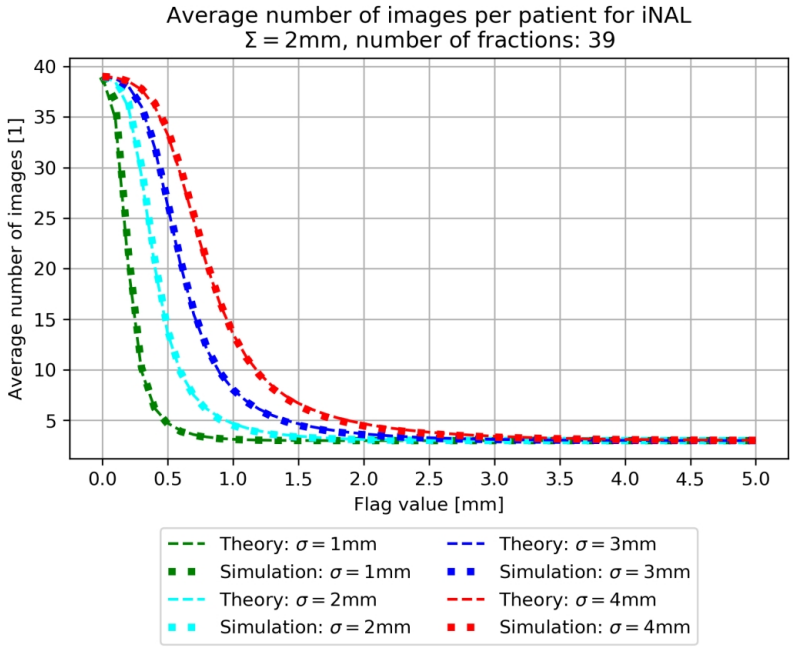
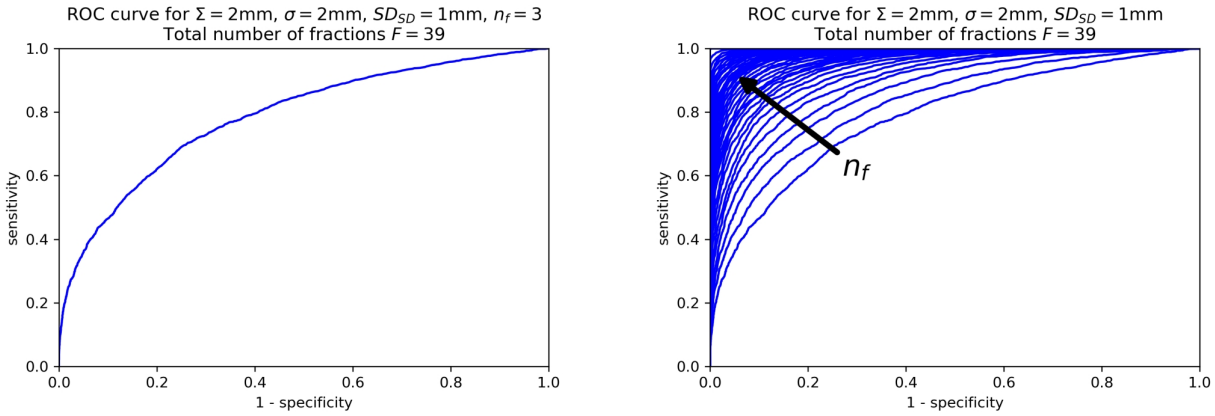
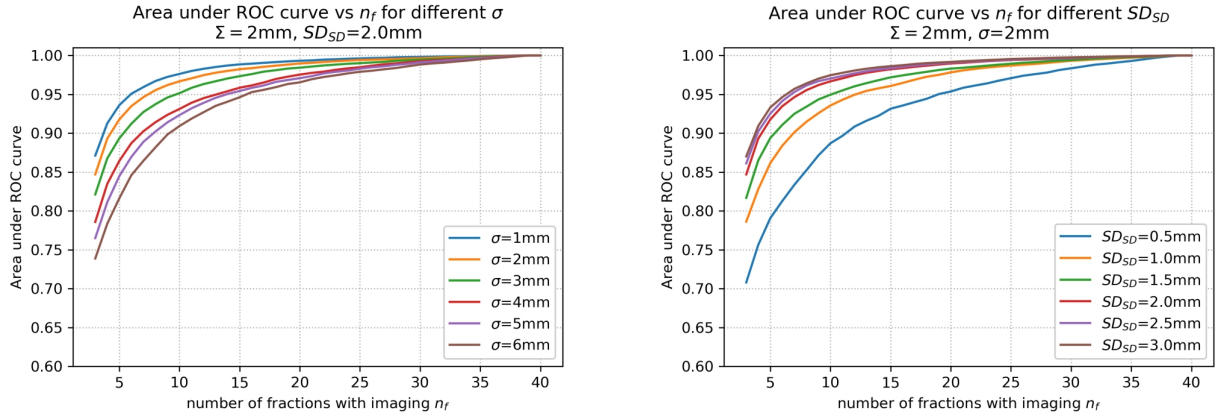


Figure 6.9: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Width of Gaussian distribution of SD_p was taken as $SD_{SD} = 0.26 \cdot \sigma$. Calculations done for 39 fractions treatment.



(a) (b)

Figure 6.10: The ROC curves achieved for $\Sigma = 2$ mm, $\sigma = 2$ mm and $SD_{SD} = 1.0$ mm. Left image shows ROC where the classifier was $SEM_{\overline{m_p}}(n_f = 3)$. Right image shows ROC curves for different classifiers: $SEM_{\overline{m_p}}(n_f)$, arrow indicates the direction of increasing n_f.



(a) (b)

Figure 6.11: AUC dependence on number of fractions with imaging — n_f, for $\Sigma = 2$ mm. Dependencies on σ (left figure) and SD_{SD} (right figure) are presented.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.76	0.73	0.68	0.70	0.68	0.68
1.0	0.82	0.80	0.76	0.72	0.70	0.69
1.5	0.85	0.83	0.81	0.78	0.74	0.73
2.0	0.87	0.84	0.82	0.79	0.78	0.75
2.5	0.89	0.87	0.85	0.83	0.79	0.77
3.0	0.90	0.87	0.86	0.83	0.81	0.79

Table 6.1: Values of AUC for $n_f = 3$, $\Sigma = 1$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.77	0.73	0.71	0.68	0.64	0.67
1.0	0.82	0.79	0.76	0.74	0.72	0.69
1.5	0.85	0.82	0.80	0.77	0.76	0.72
2.0	0.87	0.85	0.81	0.80	0.78	0.75
2.5	0.89	0.86	0.84	0.82	0.80	0.79
3.0	0.90	0.87	0.85	0.83	0.82	0.79

Table 6.2: Values of AUC for $n_f = 3$, $\Sigma = 2$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.77	0.74	0.68	0.69	0.68	0.67
1.0	0.82	0.79	0.77	0.72	0.72	0.70
1.5	0.85	0.82	0.80	0.76	0.73	0.71
2.0	0.87	0.85	0.82	0.80	0.78	0.76
2.5	0.89	0.87	0.85	0.82	0.79	0.76
3.0	0.91	0.88	0.86	0.83	0.82	0.80

Table 6.3: Values of AUC for $n_f = 3$. $\Sigma = 3$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 25$. Values of σ and SD_{SD} are given in mm.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.76	0.71	0.71	0.69	0.67	0.66
1.0	0.81	0.78	0.75	0.72	0.70	0.69
1.5	0.84	0.82	0.79	0.77	0.72	0.73
2.0	0.87	0.85	0.81	0.80	0.76	0.73
2.5	0.88	0.86	0.83	0.82	0.79	0.77
3.0	0.90	0.87	0.85	0.84	0.81	0.78

Table 6.4: Values of AUC for $n_f = 3$, $\Sigma = 1$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.76	0.71	0.68	0.70	0.67	0.65
1.0	0.82	0.79	0.74	0.74	0.72	0.69
1.5	0.84	0.82	0.79	0.77	0.72	0.72
2.0	0.87	0.85	0.82	0.79	0.76	0.74
2.5	0.89	0.86	0.84	0.81	0.80	0.77
3.0	0.91	0.87	0.85	0.83	0.80	0.78

Table 6.5: Values of AUC for $n_f = 3$, $\Sigma = 2$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm.

SD_{SD} \ σ	1.0	2.0	3.0	4.0	5.0	6.0
0.5	0.75	0.71	0.69	0.66	0.67	0.62
1.0	0.81	0.78	0.75	0.71	0.71	0.66
1.5	0.85	0.81	0.79	0.75	0.73	0.69
2.0	0.87	0.84	0.81	0.79	0.77	0.74
2.5	0.89	0.86	0.84	0.81	0.79	0.76
3.0	0.90	0.87	0.85	0.83	0.82	0.79

Table 6.6: Values of AUC for $n_f = 3$. $\Sigma = 3$ mm and different σ and SD_{SD}. Total number of fractions for this calculations was $F = 39$. Values of σ and SD_{SD} are given in mm.

6.3.3 Evaluation and comparison of protocols

Comparison of Σ_{res} was done for the simulated MC patient population as well as for the population of large movers. In Fig. 6.12-6.13 it can be seen that for a small average number of fractions with images $N \leq 5$ the iNAL protocol has similar Σ_{res}, compared to NAL and CA, for the whole population of patients. At the same time a decrease in Σ_{res} is observed with iNAL for large movers. Comparison of distributions of residual systematic errors for the whole population and large movers for CA and iNAL protocols is shown in Fig. 6.14-6.15. It can be seen that for both protocols distributions of $\overline{m_{p,i}}$ for the whole population of patients are similar. Contrary to that for large movers distributions of $\overline{m_{p,i}}$ achieved for both protocols are different in shape. For iNAL protocol it can be seen that more patients has smaller residual systematic error with simultaneous decrease in number of patients having big residual systematic error. Even bigger effect can be observed while comparing iNAL with eNAL where bigger average number of fractions with imaging is used (see Fig. 6.16-6.17). Additionally, all presented histograms show a change in shape in the distribution of setup errors after implementation of verification protocols. The distribution which initially was Gaussian was changed, and in consequence an application of van Herk margin leads to bigger percentage of patients with CTV inside margin (i.e. with residual systematic setup errors inside margin) than assumed (90%). That effect is related to the fact that 2.5 factor in van Herk margin recipe (2.14) comes from analysis of 3D Gaussians.

Fig. 6.18-6.21 show a comparison of OVP for different σ and SD_{SD}. The percentage of patients with residual systematic setup error inside the margin is shown. It can be seen that with σ increasing the percentage of large movers with CTV inside the margin also increases. With SD_{SD} increasing the percentage decreases. The difference between the protocols is biggest for the biggest SD_{SD}. The bigger is the σ value the difference is slightly smaller.

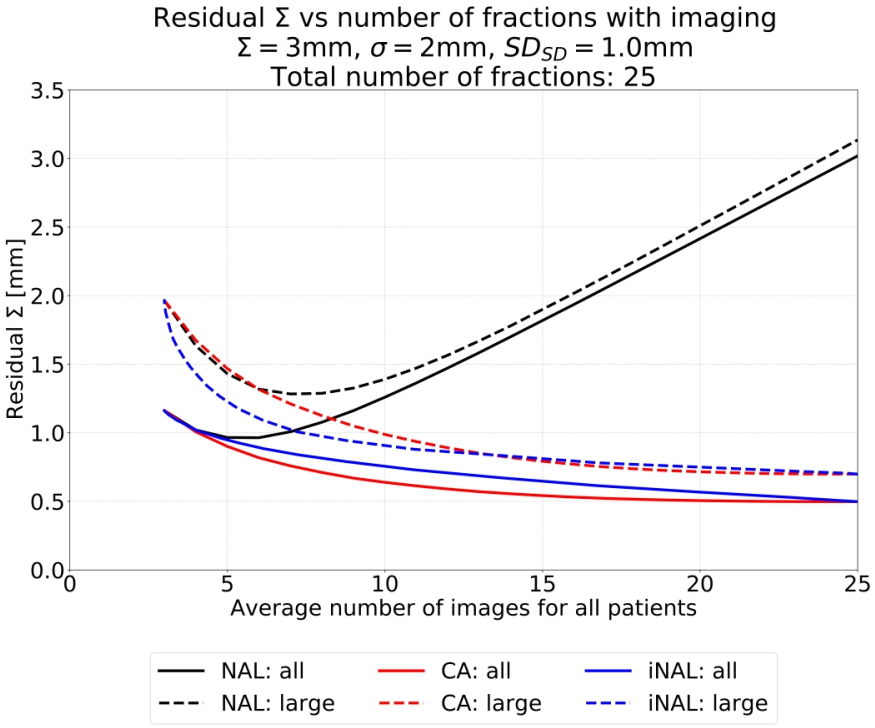


Figure 6.12: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment.

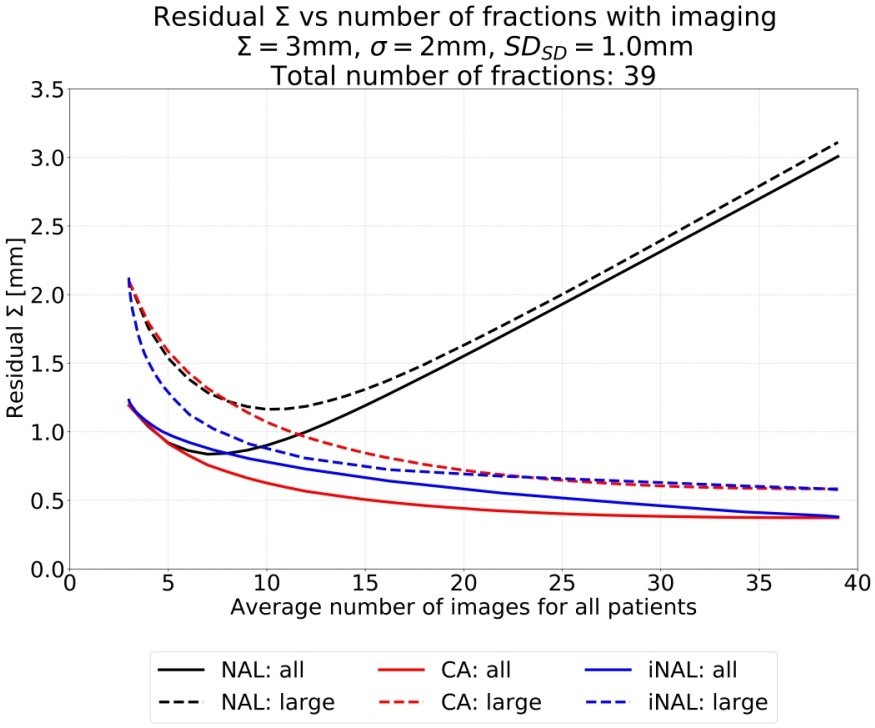


Figure 6.13: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment.

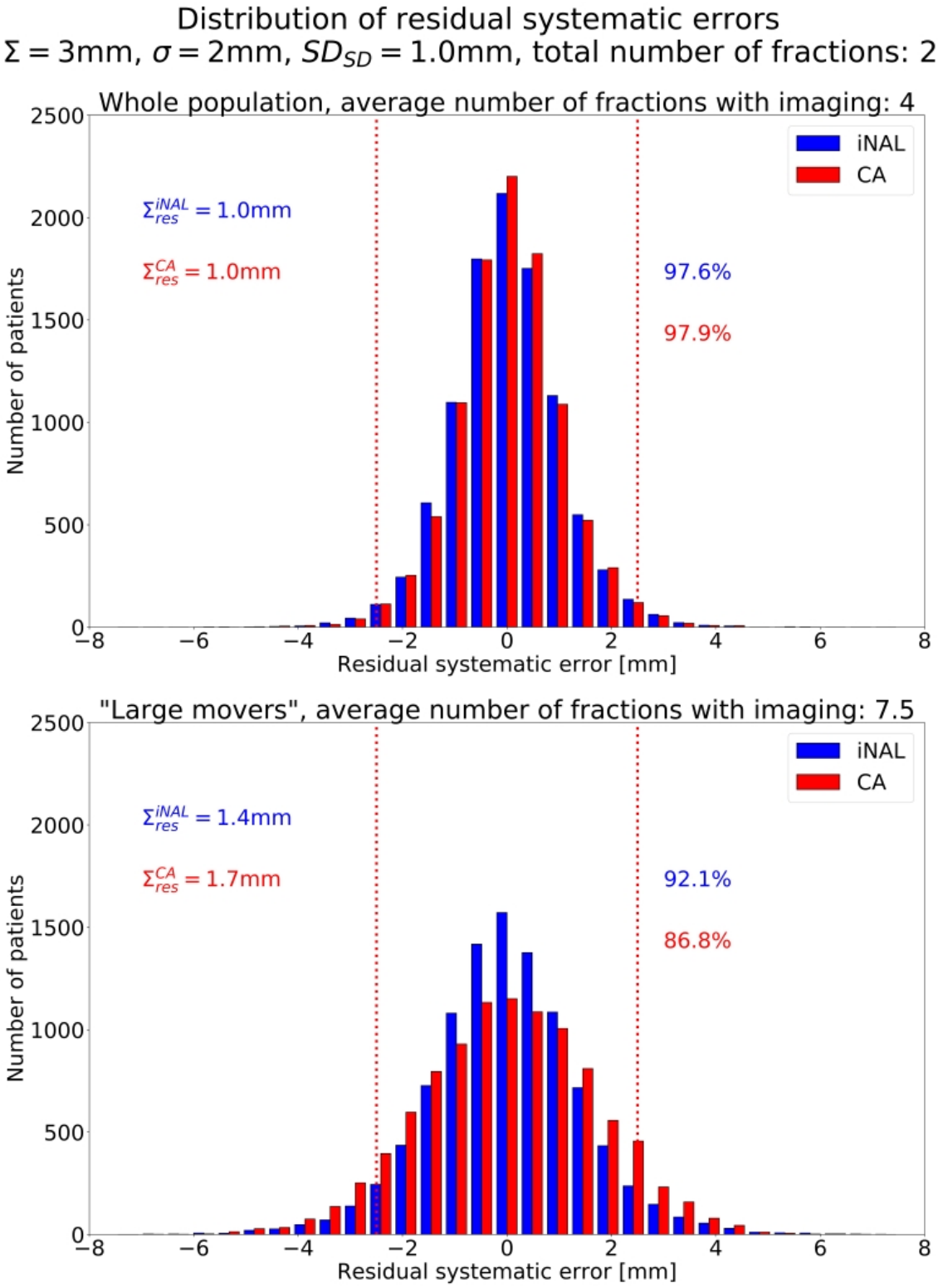


Figure 6.14: Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: CA and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 25 fractions treatment.

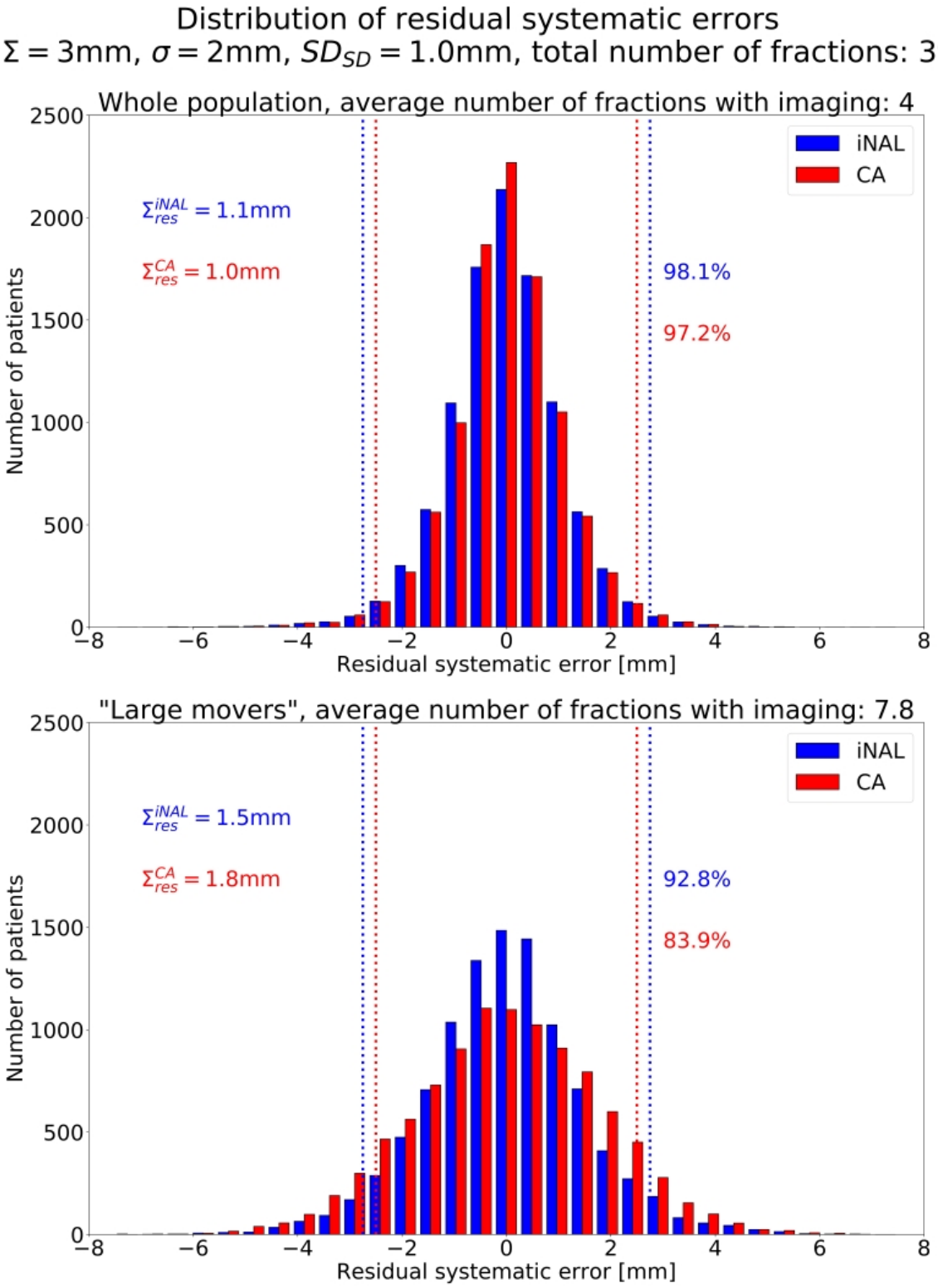


Figure 6.15: Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: CA and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 39 fractions treatment.

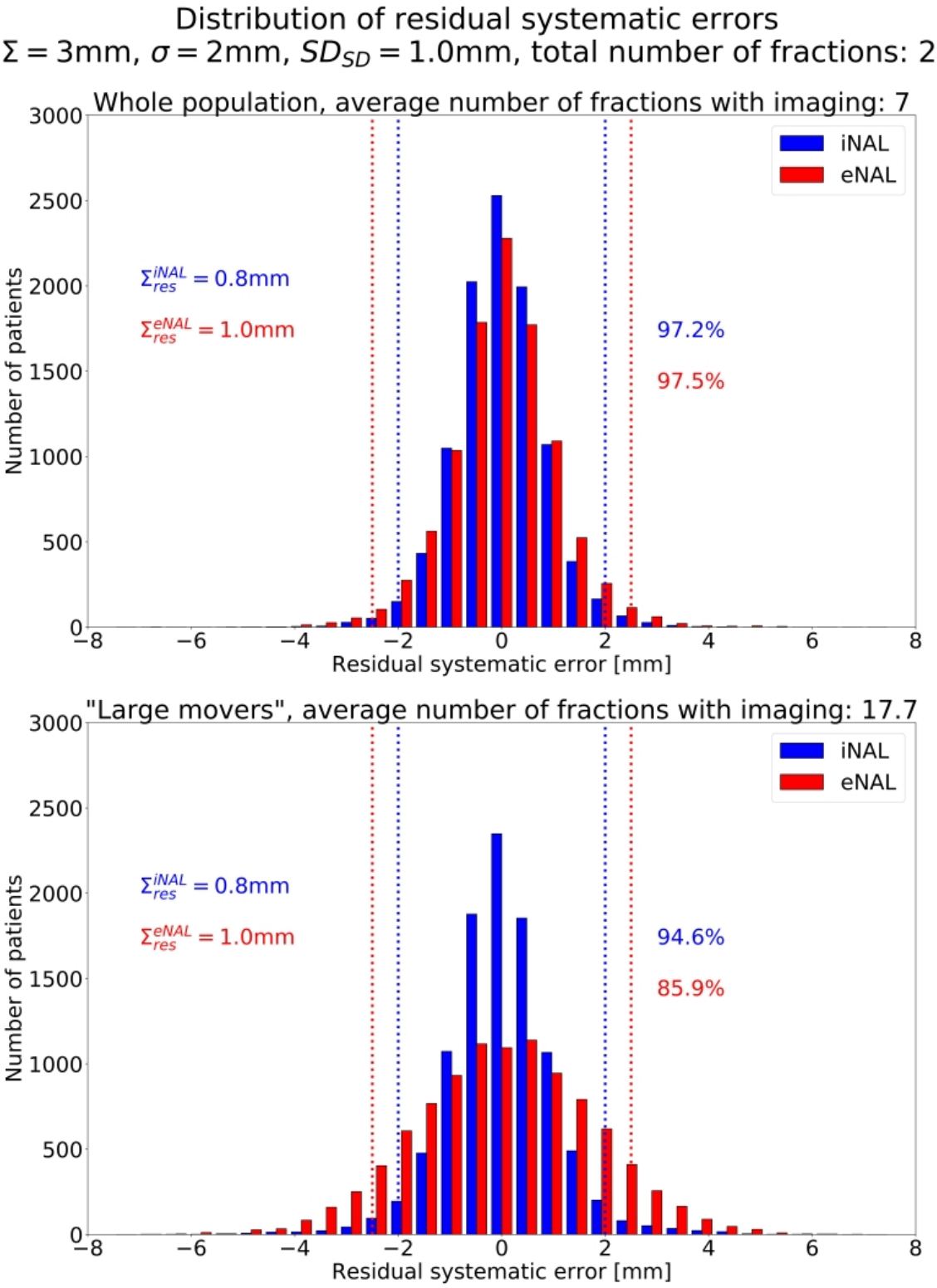


Figure 6.16: Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: eNAL and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 25 fractions treatment.

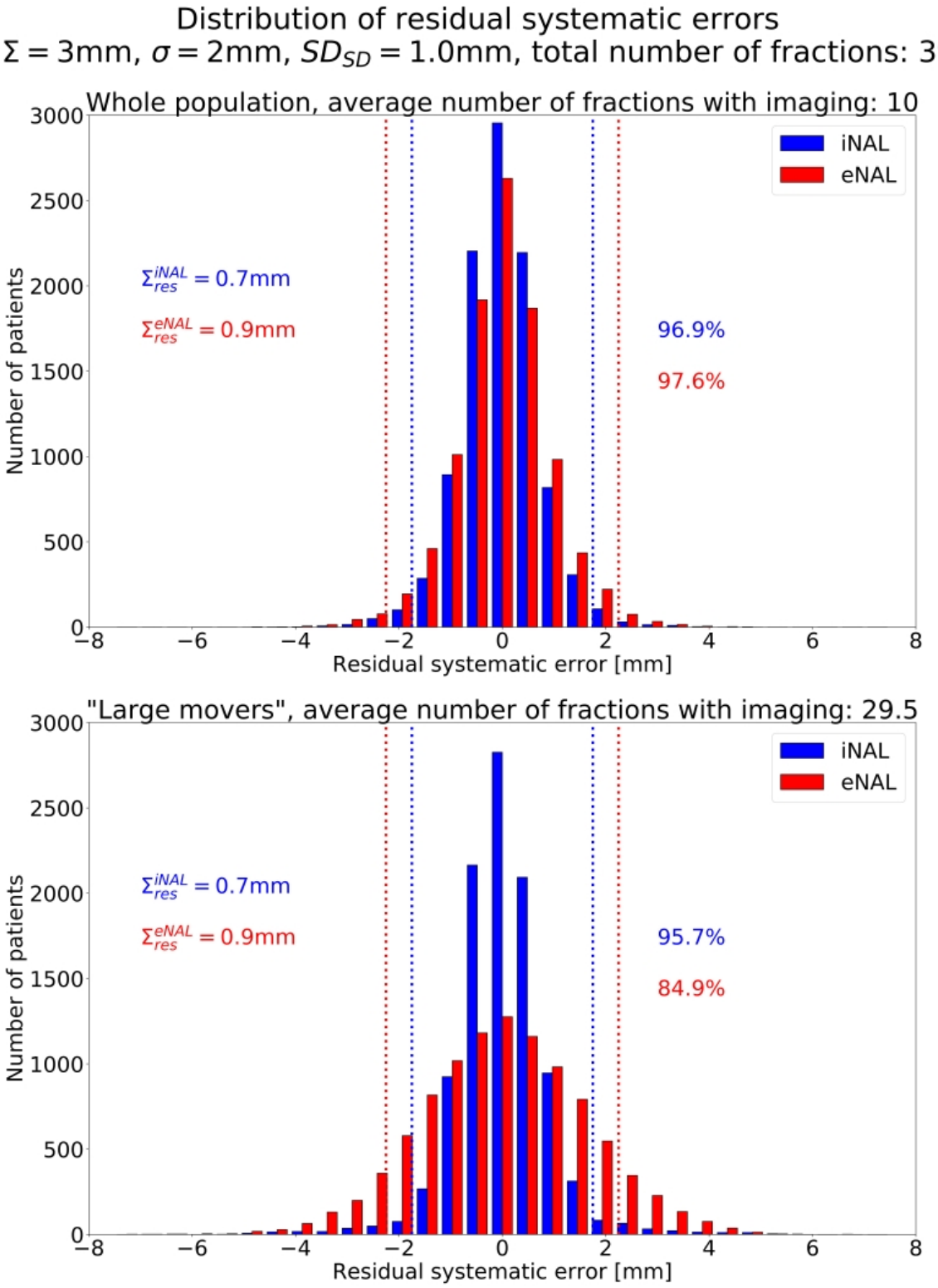


Figure 6.17: Comparison of residual systematic setup errors $\overline{m_p}$ distributions achieved with two different OVP: eNAL and iNAL. Distributions for the whole population (upper figure) and large movers (lower figure) are shown. Percentage of patients inside the margin calculated with van Herk recipe is shown. Margin size is presented with dotted lines as calculated for whole population. Calculations done for 39 fractions treatment.

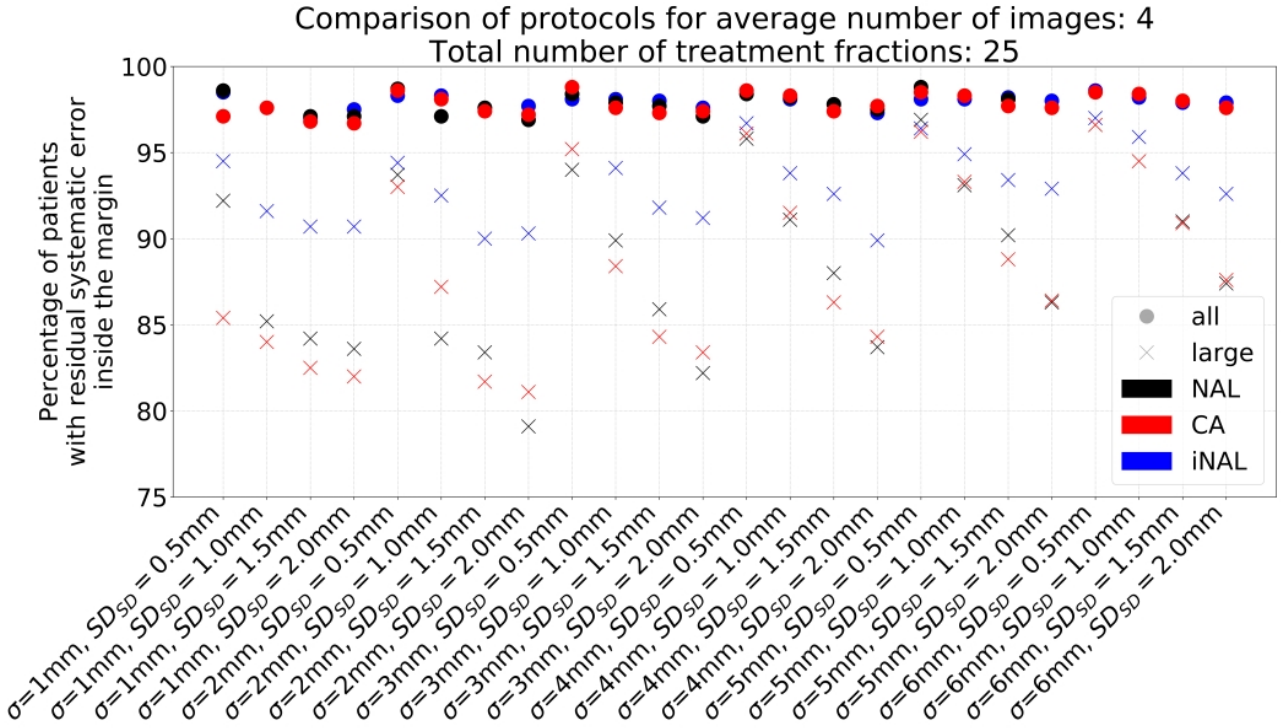


Figure 6.18: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment.

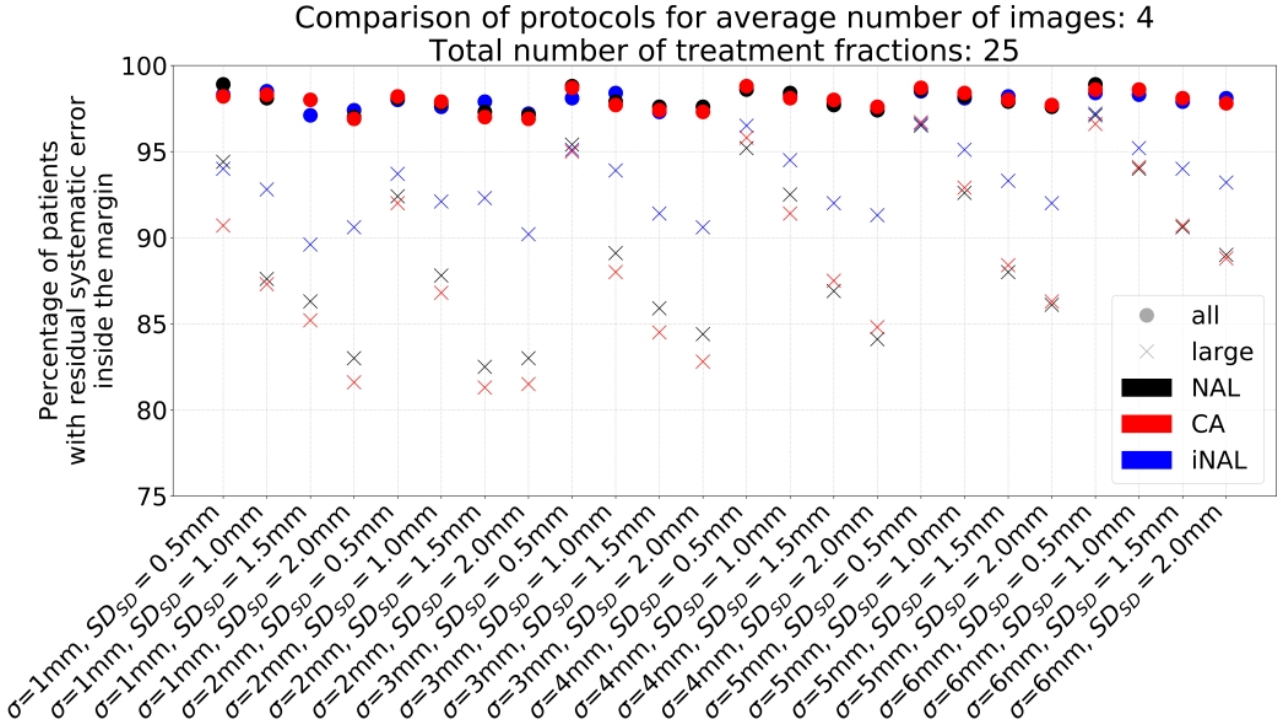


Figure 6.19: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 25 fractions treatment.

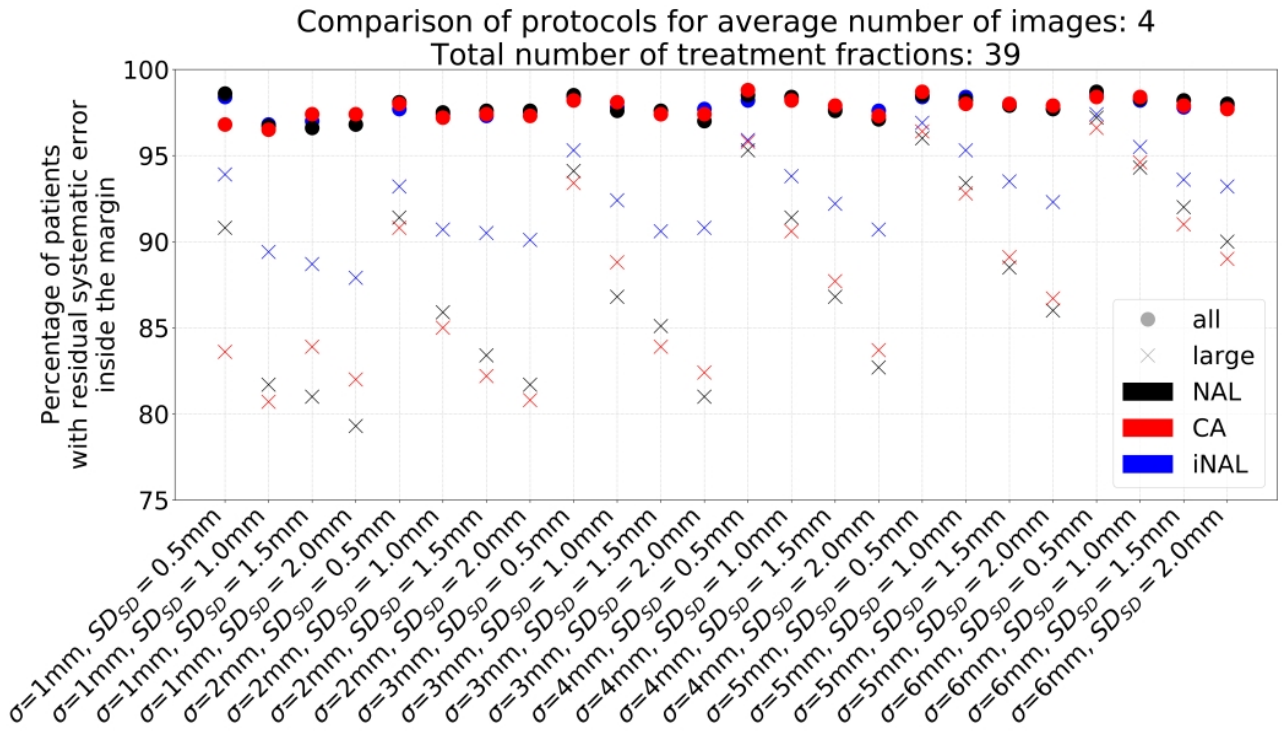


Figure 6.20: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment.

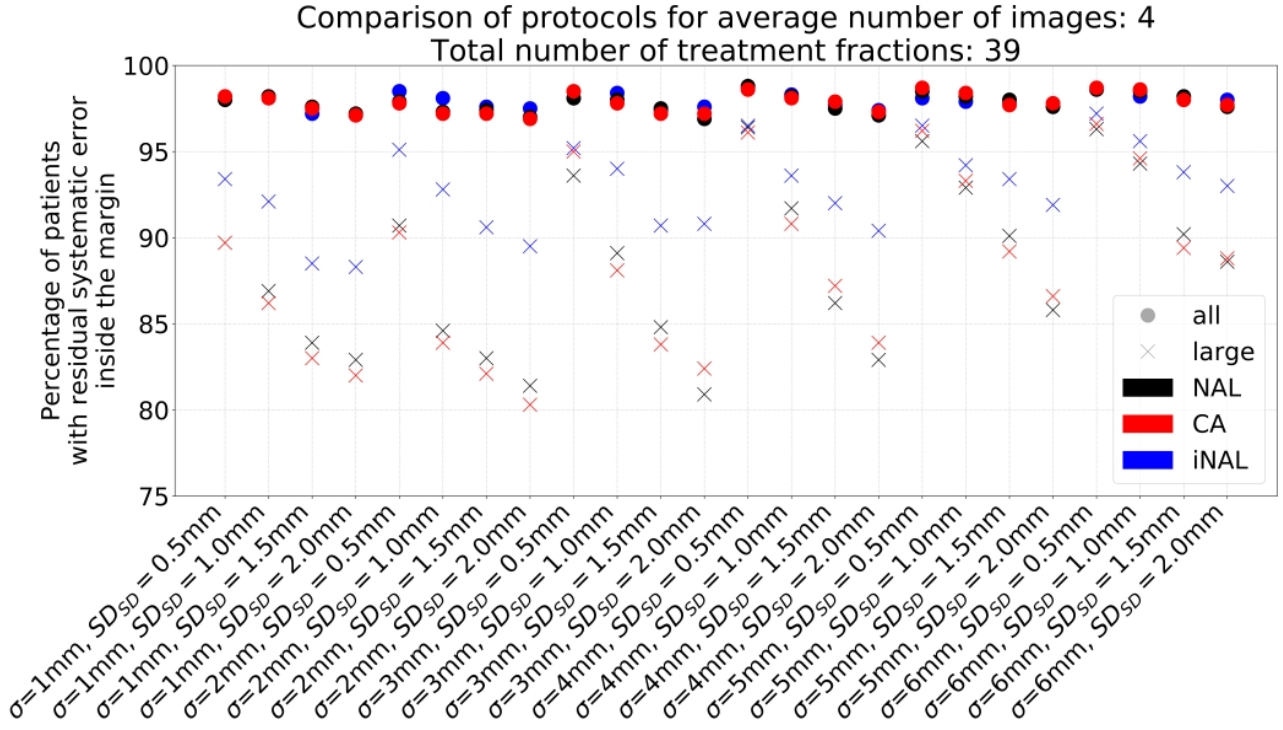


Figure 6.21: Dependence between average number of images per patient and the flag value. Theoretical results are showed with dashed lines. Simulation results are showed with dotted lines. Calculations done for 39 fractions treatment.

6.4 Application recipe

In order to provide an easy way for implementing the proposed iNAL protocol into the clinical practice a step-by-step approach is provided below. Prior to implementation of a new verification protocol it is necessary to evaluate previously treated patients coming from the same population group (i.e. receiving the same kind of treatment, the same cancer localization, the same position during radiotherapy, the same setup verification method).

1. For previously treated patients perform a statistical analysis. Check whether a significant inter-fraction time trend is found (see Section 3.6 for test description):

   (yes) Consider application of eNAL verification protocol (see Section 2.3.2)

   (no) Follow the remaining steps for application of iNAL verification protocol

2. Decide on average number of images in population which you want to use for patient verification. Check the desired flag value. In order to do that you have to know Σ, σ and SD_{SD} values for your population. You may consider using Fig. 6.8-6.9 to find the relation between average number of images in patient population and flag value. Alternatively, the theory provided in Section 6.2.4 may be used. Please keep in mind that results do not depend much on Σ.

3. Implement the iNAL protocol: create a workload in which it would be possible to calculate $SEM_{\overline{m_p}}$ and compare it with flag. On the basis of this comparison a decision shall be made to perform additional image or not.

6.5 Discussion

There are several off-line verification protocols still used in many clinics where on-line verification have not been introduced. The most popular is NAL [14] and its extension to incorporate time trends: eNAL [16]. The other popular protocol is Shrinking Action Level (SAL) [3] in which number of fractions with images depends on the action level applied to mean setup error value. The action level is decreasing with increasing number of images.

To the best of my knowledge the proposed iNAL protocol is the only one which tries to detect problematic patients i.e. those with relatively large SD of setup errors. For these patients a proper estimation of mean setup error require bigger number of images.

The proposed iNAL verification protocol allows to distribute performed images in a more clever way. With the usage of iNAL more images would be done for patients with bigger probability of being a large mover. Performance of iNAL protocol changes the distribution of $\overline{m_p}$ especially for large movers group.

Theoretical basis of the iNAL protocol has been given. It is possible to select an appropriate flag for given population of patients on the basis of χ^2 statistics. In clinical practice the balance between adequate average number of fractions with imaging and sufficient detection of large movers has to be found.

Chapter 7

CTV-PTV margins in case of time trends

7.1 Background

Generally, CTV-PTV margin recipes for coping with tumor set-up variations consider systematic and random patient set-up errors [82, 90]. As it was shown in Chapter 3 and further investigated in Chapter 4 significant inter-fraction time trends may be observed during course of treatment. To the best of my knowledge, there are no margin recipes that explicitly handle inter-fraction time trends. To cover this gap a novel margin recipe is proposed [34]. The proposed margin recipe was compared with the well-known recipe by van Herk et al. [90]. The validation of the recipe was in part based on a large database with daily measured tumor set-up errors for prostate cancer patients from Erasmus MC database.

7.2 Materials and Methods

7.2.1 Margin recipe in case of time trends

As it was shown in Chapter 4 in case of presence of time trends in order to describe setup errors we have to use mean setup error or offset value, trend slope and standard deviation of residual errors (see Fig. 7.1). In order to made the problem symmetrical a decision was made to use mean setup error which is equal to middle position of setup error (see Section 7.2.1.1).

The derivation of the proposed margin recipe is the same as the one of Marcel van Herk et al. [90] (see Section 2.2.1). At first the setup errors are divided into two groups: treatment execution (random errors) and treatment preparation errors (systematic errors). In the presented approach the time trend is considered as a specific case of systematic error around which random errors (residual ones) are distributed (see Fig. 7.1).

Requirement around which van Herk constructs his margin recipe (in case of systematic errors) is that certain percentage of patient population would receive specified dose to CTV. Marcel van Herk imposed a 90% probability that the full CTV, which would be shifted according to systematic error, would be inside a CTV-to-PTV margin and therefore would receive required dose (i.e. 95% of prescribed dose). With this requirement van Herk showed that margin size have to be 1.64Σ in case of 1D and 2.5Σ in case of 3D errors. The formulae given by van Herk ($2.5\Sigma + 0.7\sigma$) is used worldwide in clinical practice. In order to construct a 3D anisotropic margin with this formulae, one has to apply it independently for each particular direction (assuming a 3D symmetrical case — for details see Section 7.2.1.2).

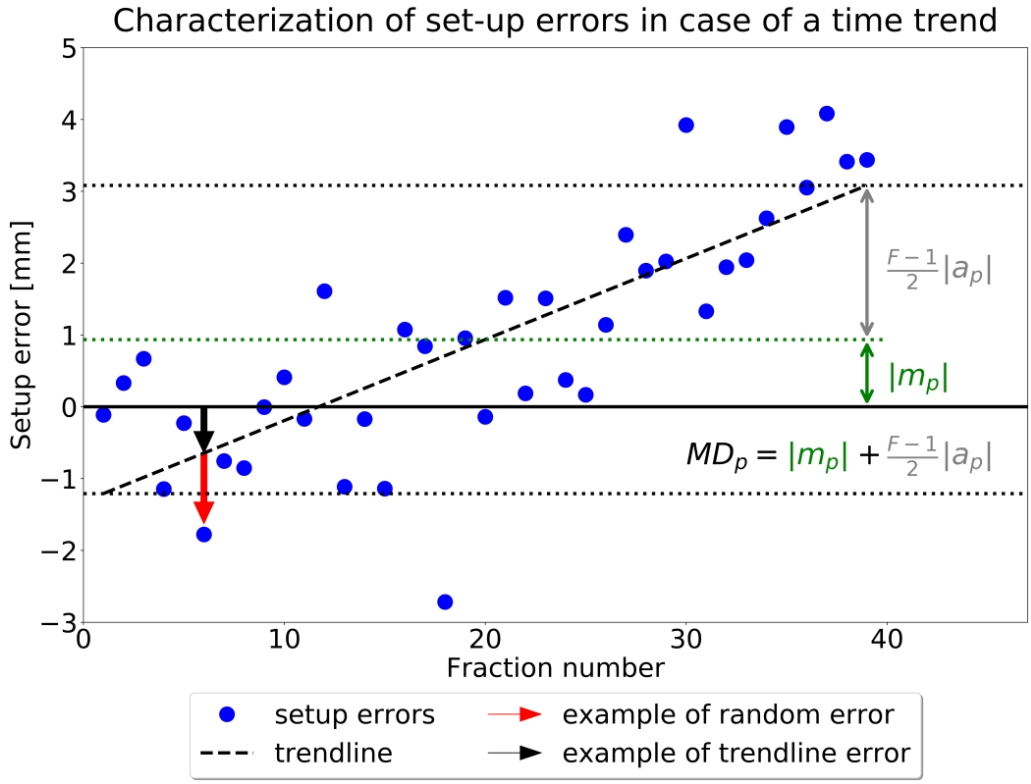


Figure 7.1: Example of setup errors for one patient. Trendline is shown. Red arrow indicates one of residual errors (i.e. errors due to trend line). Black arrow indicates one of the trendline errors (i.e. difference between zero position and trendline). Illustration of the concept of maximum deviation is also provided: green arrow shows middle position value, while grey one shows trend change for half of treatment.

The proposed novel recipe is based of a similar procedure. The distribution of CTV positions (including time trends) for population of patients is considered and requirement that 90% (or other percentage) of patients would receive 95% of dose is imposed. That gives margin component relative to the systematic error. The margin component relative to random errors is assumed to be exactly the same as in van Herk formulae. Because of that, in case of no time trend the proposed recipe simplifies to the van Herk's one.

In order to calculate the margin component for systematic errors a quantity called maximum deviation is introduced (see Fig. 7.1). Maximum Deviation (MD) describes the maximum absolute deviation from CT (according to trend) within each all systematic setup errors for a particular patient would lie. The time trend can be expressed with middle position as:

$$x_{p,i}(f) = \frac{f - (F+1)}{2} \cdot a_{p,i} + \overline{m_{p,i}} \tag{7.1}$$

where:

$p \rightarrow$ patient

$i \rightarrow$ direction

Therefore MD can be calculated as:

$$MD = |\overline{m_{p,i}}| + \frac{(F-1)}{2} \cdot |a_{p,i}| \tag{7.2}$$

Knowing the distribution of $\overline{m_{p,i}}$ and $a_{p,i}$ the maximum deviation can be calculated by a simulation. In order to do that the middle position as well as the slope have to be randomized many times and for each randomization MD_i has to be calculated. That provides the distribution of MD_i. In case of 1D calculation the margin size would be such a value of MD_i that would correspond to 90th percentile of its distribution.

In case of 3D similar approach as in van Herk recipe is used (see 7.2.1.2). The idea is to calculate MD_i for each direction as it would be a 3D case with the same distributions of $\overline{m_{p,i}}$ and $a_{p,i}$ in all axes: x, y and z. Simulations for each of x, y and z are done separately. Afterwards the radius of MD_i is calculated:

$$MD_{radius} = \sqrt{(MD_x)^2 + (MD_y)^2 + (MD_z)^2} \tag{7.3}$$

The 3D margin for axis i is calculated as 90th percentile of MD_{radius} distribution for that axis. The Python code which was used in order to calculate 3D margin is presented below.

```
import numpy as np # import numpy package

def TimeTrend3Dmargin(F, mean_m, std_m, mean_a, std_a, percentile, N = 10**7
                                        ):
  # F - number of fractions
  # mean_m - mean middle position for population of patients (in one
                                        direction)
  # std_m - std of middle position for population of patients (in one
                                        direction)
  # mean_a - mean slope position for population of patients (in one
                                        direction)
  # std_a - std of middle position for population of patients (in one
                                        direction)
  # N âĂŞ number of samples used for simulation

  # percentile - imposed percentage of patients who should receive 95\% of
                                        dose in CTV

  # Calculate distribution of maximum deviation (MD) 3 times with same
                                        parameters
  # Calculation is done by simulation each time in 1D
  mX = np.random.normal(loc=mean_m,scale=std_m,size=N)
  aX = np.random.normal(loc=mean_a,scale=std_a,size=N)
  MD_X = np.abs(mX) + (F-1)*np.abs(aX)/2

  mY = np.random.normal(loc=mean_m,scale=std_m,size=N)
  aY = np.random.normal(loc=mean_a,scale=std_a,size=N)
  MD_Y = np.abs(mY) + (F-1)*np.abs(aY)/2

  mZ = np.random.normal(loc=mean_m,scale=std_m,size=N)
  aZ = np.random.normal(loc=mean_a,scale=std_a,size=N)
  MD_Z = np.abs(mZ) + (F-1)*np.abs(aZ)/2

  # Calculate the radius of 3D distribution
```

```
  MD_3D = np.vstack((MD_X,MD_Y,MD_Z)).T
  MD_radius = np.sqrt(MD_3D[:,0]**2+MD_3D[:,1]**2+MD_3D[:,2]**2)

  # return the margin size in one direction
return np.percentile(MD_radius, percentile)
```

It is important to notice that the mean slope value does not influence the calculated margin. It can be explained by the fact that the middle position is used in MD concept. Even in case of all positive slope values MD would remain the same — the middle position depends on absolute slope value no matter the sign is. In case the mean middle position would not be equal zero there is a systematic error for the whole population. That error should be eliminated or the margin should be calculated for mean middle position equal to zero and shifted afterwards by the vector: $-\overline{\mathbf{M}}$ (see Eq. 2.5). That solution is the same as for van Herk recipe.

7.2.1.1 Proof that mean setup error is equal to the middle position

Let denote that:

$r_f \rightarrow$ random error in fraction f (residual error)

$x(f) = f \cdot a + b \rightarrow$ time trend

$m_f = f \cdot a + b + r_f \rightarrow$ real setup errors (real positions)

$F \rightarrow$ total number of fractions

Middle position would be given as:

$$mid = \frac{(a+b)+(F \cdot a+b)}{2} = \frac{(F+1)}{2} \cdot a + b \tag{7.4}$$

Mean position i.e. mean setup error can be calculated as:

$$\overline{m_f} = \frac{1}{F}\sum_{f=1}^{F}(f \cdot a + b + r_f) = \frac{1}{F} \cdot \left[\sum_{f=1}^{F}(f \cdot a + b) + \sum_{f=1}^{F} r_f\right] \tag{7.5}$$

According to the fact that the time trend is defined with OLS, for residuals errors it is true that $\sum_{f=1}^{F} r_f = 0$. Therefore:

$$\overline{m_f} = \frac{1}{F}\sum_{f=1}^{F}(f \cdot a + b) = \frac{(F+1)}{2} \cdot a + b = mid \tag{7.6}$$

Table 7.1 presents comparison between mean and SD calculated directly from the distributions of $\overline{m_p}$ with those calculated as middle position and its SD. The only difference can be seen for head-feet direction and is considered to be insignificant (0.01 mm).

	Standard Parametrization		**Time Trend Parametrization**	
	M **mm**	Σ **mm**	M **mm**	Σ **mm**
left-right	-0.32	2.5	-0.32	2.5
anterior-posterior	-0.54	3.47	-0.54	3.47
head-feet	-0.97	3.37	-0.97	3.36

Table 7.1: Comparison of calculated parameters for standard and time trend parametrization. Mean position of standard parametrization is the same as middle position of time trend parametrization.

7.2.1.2 3D margin calculation

Assume that systematic error is described in three principal directions with normal distributions: $G(0, \Sigma_x)$, $G(0, \Sigma_y)$, $G(0, \Sigma_z)$. In order to calculate margin $\mathrm{M}_x^{\mathrm{PTV}}$ in x direction one has to find such $\mathrm{M}_x^{\mathrm{PTV}}$ that probability of the setup error within the margin:

$$P\left((x_i)^2/(\mathrm{M}_x^{\mathrm{PTV}})^2 + (x_j)^2/(\mathrm{M}_x^{\mathrm{PTV}})^2 + (x_k)^2/(\mathrm{M}_x^{\mathrm{PTV}})^2 \leq 1\right)$$

would be equal to 90%, with x_i, x_j, x_k randomized independently from $G(0, \Sigma_x)$. As $(x_i)^2/(\Sigma_x)^2$ has chi-squared distribution with one degree of freedom, the sum $(x_i)^2/(\Sigma_x)^2 + (x_j)^2/(\Sigma_x)^2 + (x_k)^2/(\Sigma_x)^2$ has chi-squared distribution with three degrees of freedom. Therefore discussed probability can be rewritten as:

$$P\left(\frac{(x_i)^2 + (x_j)^2 + (x_k)^2}{(\mathrm{M}_x^{\mathrm{PTV}})^2} \leq 1\right)$$

$$P\left(\frac{(x_i)^2 + (x_j)^2 + (x_k)^2}{(\Sigma_x)^2} \leq \frac{(\mathrm{M}_x^{\mathrm{PTV}})^2}{(\Sigma_x)^2}\right)$$

$$P\left((x_i)^2/(\Sigma_x)^2 + (x_j)^2/(\Sigma_x)^2 + (x_k)^2/(\Sigma_x)^2 \leq \frac{(\mathrm{M}_x^{\mathrm{PTV}})^2}{(\Sigma_x)^2}\right) = \mathrm{CDF}(\chi^2_{dof} = 3) = 0.9$$

where $(\mathrm{M}_x^{\mathrm{PTV}})^2/(\Sigma_x)^2 = q_{0.9}$ stands for 0.9 quantil of chi-squared distribution with three degrees of freedom ($\chi^2_{dof} = 3$). The same applies for other two directions. For an anisotropic 3D margin the probability would be written as: $P\left(x^2/(\mathrm{M}_x^{\mathrm{PTV}})^2 + y^2/(\mathrm{M}_y^{\mathrm{PTV}})^2 + z^2/(\mathrm{M}_z^{\mathrm{PTV}})^2 \leq 1\right)$ which can be now rewritten as:

$$P\left(x^2/(q_{0.9} \cdot \Sigma_x^2) + y^2/(q_{0.9} \cdot \Sigma_y^2) + z^2/(q_{0.9} \cdot \Sigma_z^2) \leq 1\right)$$

$$P\left(x^2/(\Sigma_x)^2 + y^2/(\Sigma_y)^2 + z^2/(\Sigma_z)^2 \leq q_{0.9}\right)$$

As all $(x^2/(\Sigma_x)^2)$, $(y^2/(\Sigma_y)^2)$ and $(z^2/(\Sigma_z)^2)$ have χ^2 distributions with one degree of freedom, their sum has a χ^2 distribution with three degrees of freedom. Therefore probability can be written as $P\left(x^2/(\Sigma_x)^2 + y^2/(\Sigma_y)^2 + z^2/(\Sigma_z)^2 \leq q_{0.9}\right) = P(\chi^2_{dof=3} \leq 0.9)$ which is equal to 0.9.

7.2.2 Fitting equation for time trend margin calculation

The proposed CTV-PTV margin can be calculated with usage of the Python code as presented in Section 7.2.1. Alternatively, an analytical equation was searched. Equation (7.2) depends

on total number of fractions F. For practical reasons that dependence should be removed. In order to do that a variable A is introduced, representing the slope in one-half of the treatment fractions. Distribution of A was described by M_A and Σ_A per direction i.

$$A_p = \frac{(F-1)}{2} \cdot a_p \tag{7.7}$$

$$M_{A,i} = \frac{(F-1)}{2} \cdot M_{a,i} \tag{7.8}$$

$$\Sigma_{A,i} = \frac{(F-1)}{2} \cdot \Sigma_{a,i} \tag{7.9}$$

After such a substitution (7.2) comes to:

$$MD = |\overline{m_{p,i}}| + |A_{p,i}| \tag{7.10}$$

Equation (7.10) is symmetrical in behaviour for $\overline{m_{p,i}}$ and $A_{p,i}$. That makes it easier to interpret and use in further margin evaluation.

A fit was done with least squares method as implemented in scipy package in case $M = 0$ mm and $M_a = 0$ mm/fraction. A wide range of $\Sigma = [0, 5]$ mm and $\Sigma_A = [0, 5]$ mm values were used giving in total 441 points for each fit. It was assumed that fitted equation should simplify to van Herk recipe in case of $\Sigma_A = 0$ mm. The fit was evaluated in terms of difference between values calculated with the Python code (Section 7.2.1) and those achieved with the proposed equation.

7.2.3 Synthetic populations

Synthetic populations were used to compare the proposed margin recipe (Section 7.2.1) with the one of van Herk (Section 2.2.1). A general description of synthetic populations was given in Section 4.2.1. All synthetic populations used at this step consist of 10^4 patients with 39 fractions treatment.

In order to perform 3D margin tests two types of synthetic populations were prepared. At first, populations with isotropic setup errors distributions were created. That means that a population distribution in each principal axis was the same. Parameters taken to creation of these populations were: $\Sigma \in \{1, 2, 3, 4\}$ mm, $\sigma \in \{1, 2, 3\}$ mm, $\sigma_\sigma = 0.75$ mm, $\Sigma_a \in \{0.0, 0.05, 0.1, 0.15\}$ mm/fraction and $M_a \in \{0.0, -0.05, -0.1\}$ mm/fraction. In total 144 isotropic populations were generated.

Secondly anisotropic populations were created. At first populations with setup error only in 1D were created with the same parameters used as for isotropic populations generation. Afterwards 3D populations were created by randomly choosing (with replacements) populations with randomized parameters. It was forced that each set of parameters would be present in at least one anisotropic population. In total 144 anisotropic populations were generated.

7.2.4 Clinical Data

In order to test the proposed margin recipe for clinical data, setup errors gathered for 835 prostate cancer patients from Erasmus MC hospital were used. For detailed description of the data see Section 3.2.2. Margins were calculated with assumption that no verification protocol is used (so no setup correction was applied).

7.2.5 Validation of margin recipe – Ellipsoid test

Validation of margin recipe would be based on determination of percentage of patients with systematic setup error not inside the margin. According to the assumption made during derivation of margin recipe it should be 10% if 90% of population is required to have CTV within prescribed dose. For each patient and each fraction setup error can be described as $[x_{p,f}, y_{p,f}, z_{p,f}]$. If that setup error lies within the margin specified as $[M_x, M_y, M_z]$ true would be:

$$\frac{(x_{p,f})^2}{(M_x)^2} + \frac{(y_{p,f})^2}{(M_y)^2} + \frac{(z_{p,f})^2}{(M_z)^2} \leq 1 \tag{7.11}$$

In order to check performance of the described above ellipsoid test the simulation was done. For each combination of $\Sigma \in \{0.5, 1.0, \ldots, 5.0\}$ mm and $\Sigma_{a,i} \in \{0.0, 0.01, \ldots, 0.10\}$ mm/fraction setup errors in 39 fractions for 10^5 patients were simulated, with i denoting the principal axis. Mean population middle position M as well as mean slope M_a were assumed to be zero. Slope was assumed to be independent for all three directions. For each set of parameters a margin was calculated assuming 80th, 85th, 90th, 95th or 99th percentile. Afterwards the ellipsoid test was performed. Results can be seen in Fig. 7.2-7.6. The ellipsoid test applied for simulated population of patients with time trends showed that for 90th percentile derivation of margin recipe on average 9.99 ± 0.09% of patients have at least one fraction outside the margin (range: 9.61-10.44%). Summary of results for all tested values of percentiles used for margin definition are given in Table 7.2. For all tested combinations the difference between expected and achieved values did not exceed 0.7%.

80%	**85**%	**90**%	**95**%	**99**%
19.97 ± 0.13	14.98 ± 0.11	9.99 ± 0.09	5.00 ± 0.07	1.00 ± 0.03
$[19.33, 20.45]$	$[14.49, 15.43]$	$[9.61, 10.44]$	$[4.73, 5.27]$	$[0.88, 1.12]$

Table 7.2: Validation of time trend margin recipe for different percentile values used to margin size definition. For each percentile more than 13310 combinations of time trend parameters (Σ, Σ_a) were used. For every combination of parameters 10^5 patients with their setup errors were simulated. First row gives mean achieved value and its standard deviation. Second row provides range of achieved values.

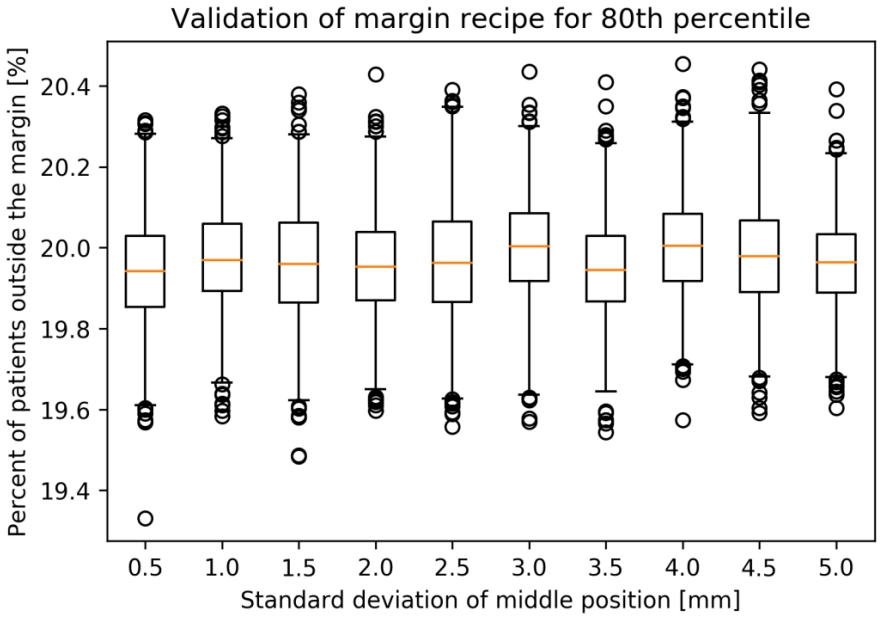


Figure 7.2: Distribution of percentage of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients.

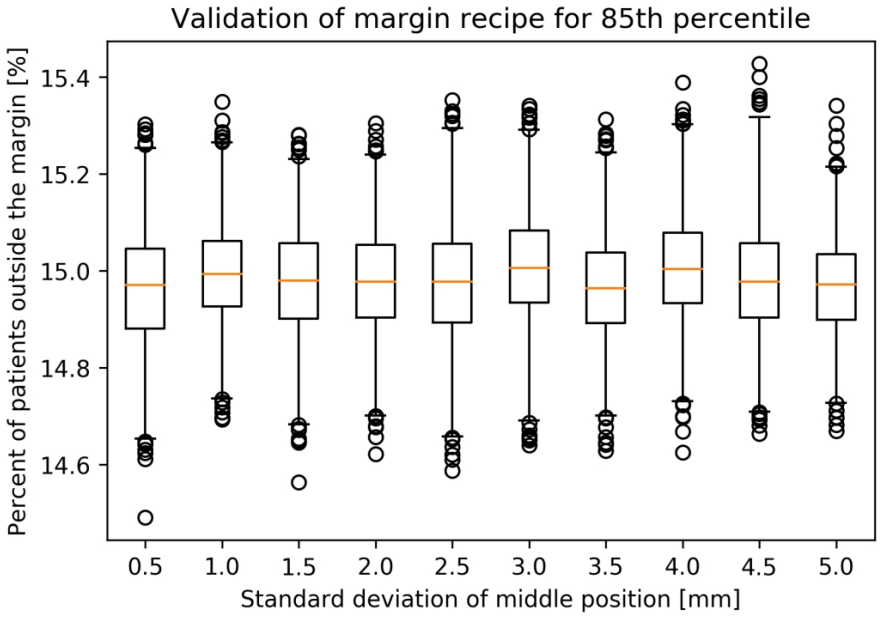


Figure 7.3: Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients.

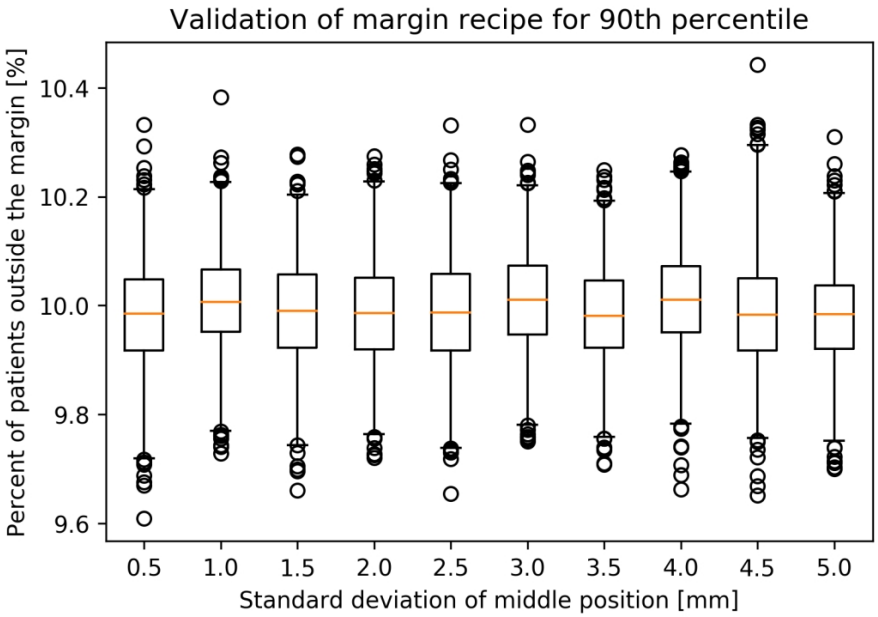


Figure 7.4: Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients.

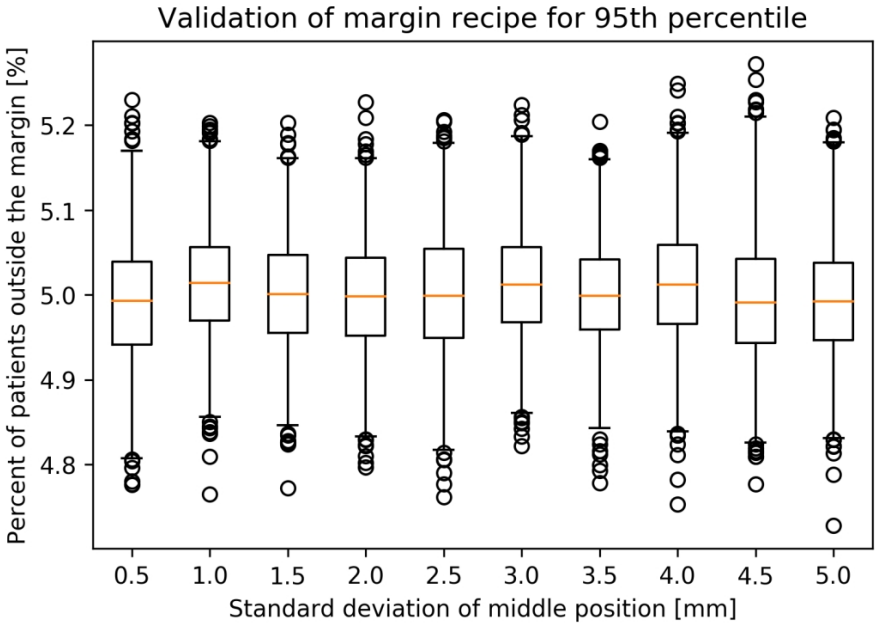


Figure 7.5: Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients.

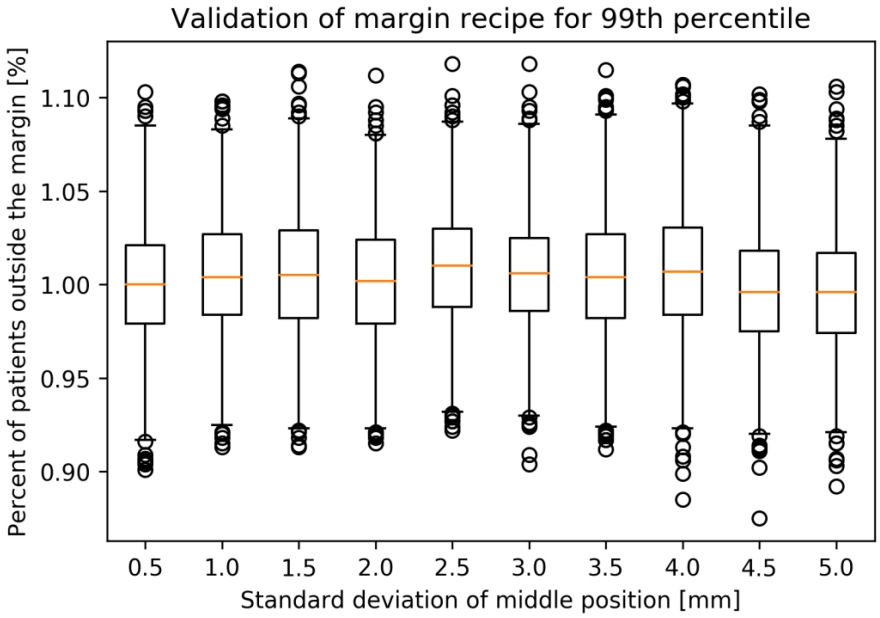


Figure 7.6: Distribution of percent of patients outside the margin achieved for different standard deviations of mean setup position. In each case 11 different standard deviations of time trend slope were used. Each set of parameters was used to simulate treatment setup errors for 10^5 patients.

7.3 Results

In case of van Herk's recipe the time trend is not considered and therefore it influences the magnitude of random errors. That is why there is a small difference between σ_i and σ_i^{tt}. In order to make fair a comparison, $0.7\sigma_i^{tt}$ was subtracted from both margins to calculate their systematic part. In other words van Herk's systematic margin was calculated as: $2.5\Sigma + 0.7(\sigma_i - \sigma_i^{tt})$.

7.3.1 Synthetic populations

Fig. 7.7 presents detailed results achieved for 144 isotropic populations. Both margin recipes (van Herk's and the proposed one) are compared in terms of percentage of patients with at least one trend error lying outside the margin. Trend error is defined here as error according to trendline as shown in Fig. 7.1. For the proposed margin recipe the percentage of patients with at least one trendline error outside the margin was $10.0\% \pm 0.2\%$. Contrary, even 100% of patients had at least one trendline error outside the margin for van Herk's recipe. With Σ increasing the margin size increases and therefore the percentage achieved for van Herk's recipe decreases. For populations without time trends (yellow squares in Fig. 7.7) the van Herk's recipe ended with 10.0% of patients with non-random error lying outside the margin. Percentage of patients outside van Herk's margin increases with increasing values of trendline parameters (M_a and Σ_a).

Results for 289 populations (among which was one clinical population) are presented in Fig. 7.8. It can be clearly seen that the percentage of patients with at least one trend setup error lying outside the CTV-PTV for van Herk's margin recipe is always bigger than for the proposed margin. Only for populations without time trend van Herk's margin recipe leads to assumed 10%. Depending on parameters the van Herk margin can be too small for the whole population (all patients outside). It has to be stressed that the percentage of patients outside the proposed margin was on average 10.0% (range [9.39,10.72]).

7.3.2 Clinical Data

The margin was calculated for Erasmus MC database using both recipes (van Herk and the proposed one).

Achieved margin sizes are presented in Table 7.3. The proposed recipe yields to bigger margin size. The difference is in order of 1.3-2 mm. The ellipsoid test described in 7.2.5 was used to calculate the percentage of patients with non-random errors lying outside the margin. Calculations were done for Erasmus MC database. The proposed recipe fulfills the assumption of 10% of patients outside the margin. On the other hand, usage of van Herk recipe will end up with 23.7% of patients with CTV outside the margin.

7.3.3 Fitting equation for time trend margin calculation

The fitted margin equation, for a principal direction i, in case when the mean slope $M_A = 0$ mm, is given below. Fitted parameters α, δ and γ depend on assumed percentile of patients who can experience non-random errors exceeding margin during the entire treatment time. Fitted values are presented in Table 7.3.3. In the same table a summary of comparison between margin

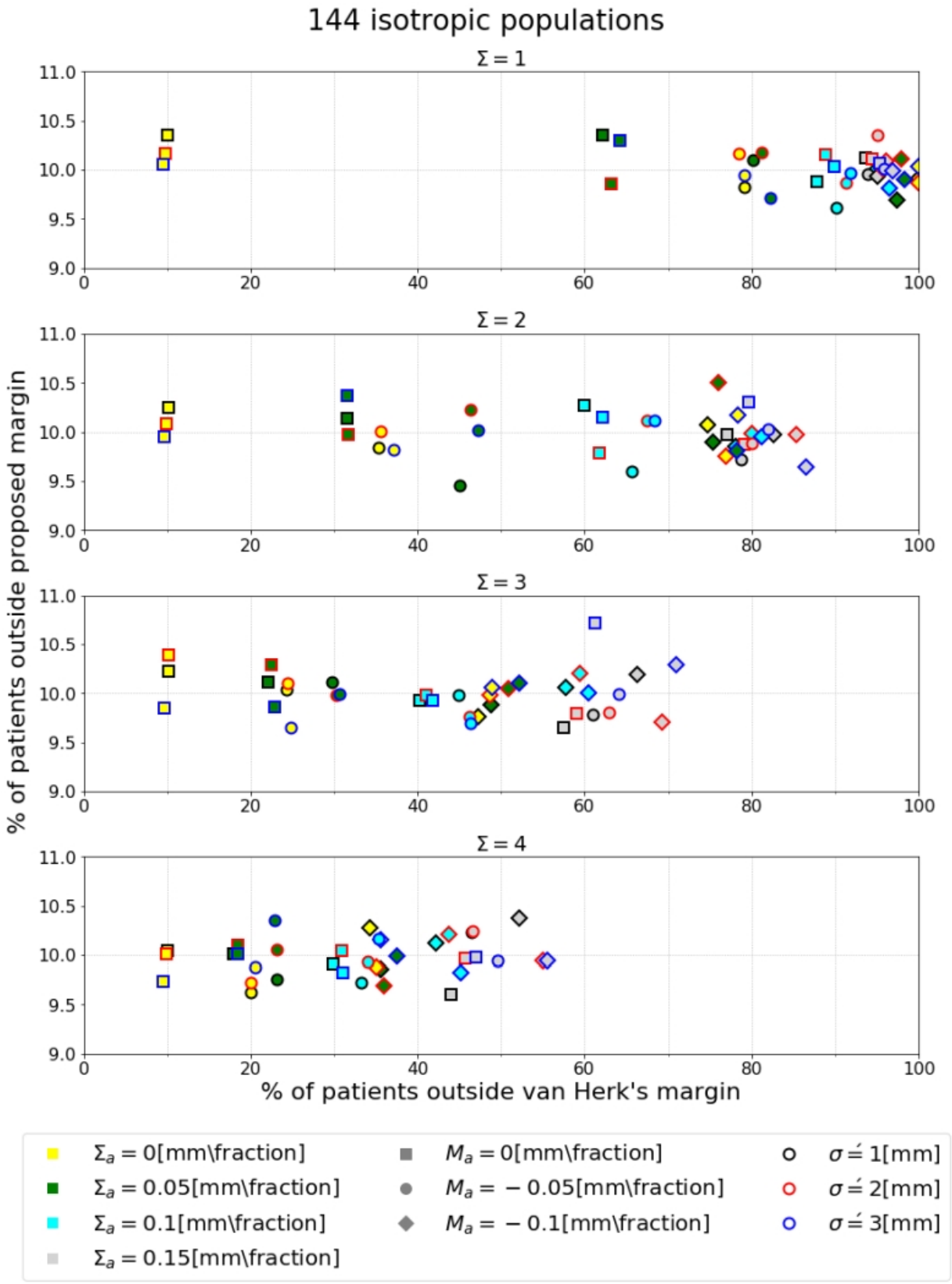


Figure 7.7: Comparison of percentage of patients with at least one trendline error lying outside the margin for van Herk and the proposed margin recipe. Results achieved for isotropic populations with different time trend simulation parameters ($M = 0$ mm, Σ, M_a, Σ_a and σ').

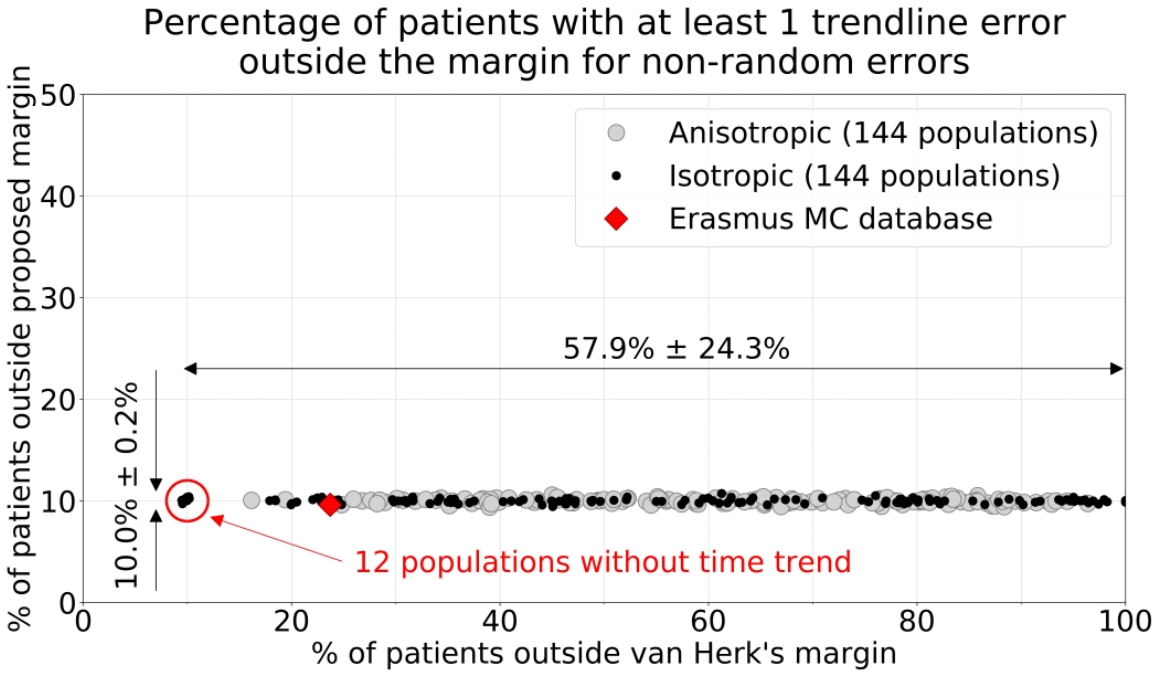


Figure 7.8: Percentage of patients with at least one trend setup error lying outside CTV-PTV 3D margin for two compared margin recipes: van Herk (x axis) and the time trend margin (y axis). In total 290 populations were analyzed.

	van Herk's Recipe non-random / total	**The proposed Recipe** non-random / total
left-right	6.3 mm / 7.6 mm	7.6 mm / 8.9 mm
head-feet	8.6 mm / 10.2 mm	10.8 mm / 12.5 mm
anterior-posterior	8.8 mm / 10.6 mm	11.0 mm / 12.8 mm
% of patients outside 3D margin	23.7%	9.6%

Table 7.3: Comparison between van Herk's and the proposed margin recipes in terms of margin size and percentage of patients with at least one trendline error lying outside the 3D margin. Values were calculated for real patient data from Erasmus MC database. Percentage of patients outside the margin was calculated with Ellipsoid Test.

percentile	α	δ	γ	Δmargin mm	α by van Herk et al.
80	2.15	-0.94	0.86	$2.65 \cdot 10^{-5}$; [-0.005, 0.004]	2.16
85	2.31	-1.08	0.67	$4.33 \cdot 10^{-5}$; [-0.006, 0.005]	2.31
90	2.50	-1.27	0.52	$8.19 \cdot 10^{-5}$; [-0.008, 0.007]	2.50
95	2.79	-1.54	0.38	$1.11 \cdot 10^{-4}$; [-0.015, 0.012]	2.79
99	3.37	-2.09	0.26	$1.75 \cdot 10^{-4}$; [-0.016, 0.019]	3.36

Table 7.4: Fit parameters for margin equation (Eq. (7.12)) as a function of percentile of patients with all non-random errors inside the CTV-PTV margin. The Δmargin expresses mean difference of comparison between margin calculation with the Python code and direct calculation using Eq. (7.12), additionally the range is given in brackets. For comparison purposes the last column contains α values as derived by van Herk et al.[90].

calculation methods (Python calculation vs direct calculation using Eq. (7.12)) is presented. More detailed comparison is shown in Fig. 7.13-7.9.

$$M^{PTV}_{trend,i}(M_i = 0, \Sigma_i, M_{A,i} = 0, \Sigma_{A,i}) = \alpha \cdot (\Sigma_i + \Sigma_{A,i}) + \frac{\delta \cdot \Sigma_i \cdot \Sigma_{A,i}}{\sqrt{\Sigma_i^2 + \gamma \cdot \Sigma_i \cdot \Sigma_{A,i} + \Sigma_{A,i}^2}} \tag{7.12}$$

It is also possible to fit the 4D equation in case $M_A \neq 0$ mm. In this case the equation become complicated with nine parameters which have to be fitted.

$$\begin{aligned} M^{PTV}_{trend,i}(M_i = 0, \Sigma_i, M_{A,i}, \Sigma_{A,i}) = \alpha \cdot (\Sigma_i + \Sigma_{A,i}) + \beta_1 \cdot |M_{A,i}| + \\ \frac{\beta_2 \cdot \Sigma_i \cdot \Sigma_{A,i}}{\sqrt{\Sigma_i^2 + \beta_3 \cdot \Sigma_i \cdot \Sigma_{A,i} + \Sigma_{A,i}^2}} + \frac{\beta_4 \cdot M_{A,i}^2}{\sqrt{\beta_5 \cdot \Sigma_i^2 + \beta_6 \cdot \Sigma_i \cdot |M_{A,i}| + M_{A,i}^2}} + \\ \frac{\beta_7 \cdot M_{A,i}^2}{\sqrt{\beta_8 \cdot \Sigma_{A,i}^2 + \beta_9 \cdot \Sigma_{A,i} \cdot |M_{A,i}| + M_{A,i}^2}} \end{aligned} \tag{7.13}$$

Results of Eq. (7.13) performance, in case of 90% assumption, are presented in Fig. 7.14-7.16. These figures present 3D plots for different values of M_A (Fig. 7.14), Σ_A (Fig. 7.15) and Σ (Fig. 7.16) as it is not possible to present 4D equation fit in one graph. The equation fitting was done for 5733 points. The fitted α parameter, which corresponds to van Herk equation, was equal 2.49. Achieved values of β parameters were: -0.12, -1.31, 0.91, 1.57, 1.61, 1.32, 0.25, 1.69 and -0.25. Mean difference between margin calculated with Python script and using Eq. (7.13) was 0.0007 (range: $[-0.1, 0.1]$). Similar results were achieved for other percentiles of patients assumed to have all trend errors lying within the CTV-PTV margin.

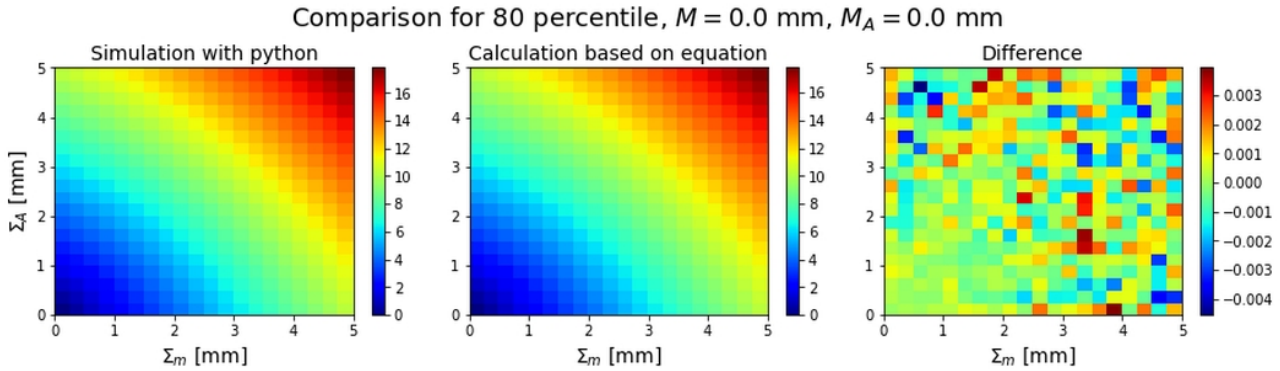


Figure 7.9: Comparison between margin calculation methods (Python code vs direct equation) for 80% of patients assumed to have all non-random error within the margin.

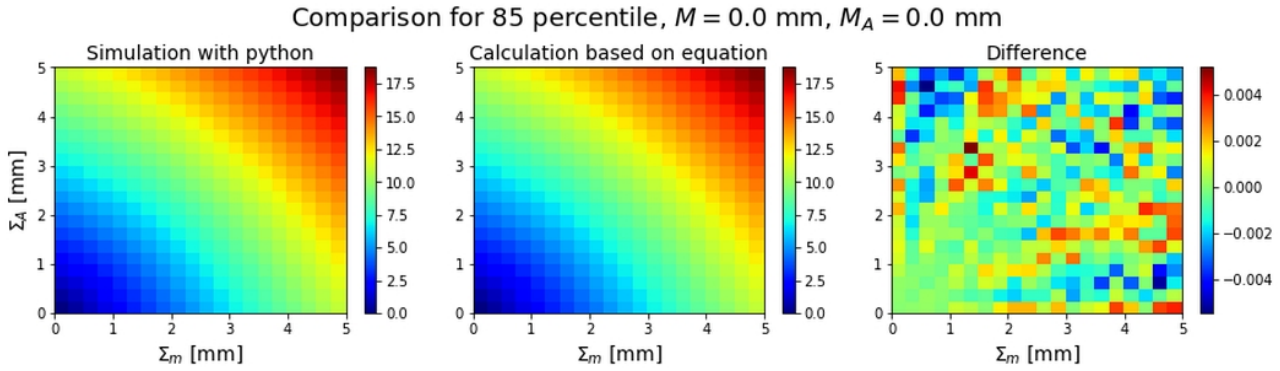


Figure 7.10: Comparison between margin calculation methods (Python code vs direct equation) for 85% of patients assumed to have all non-random error within the margin.

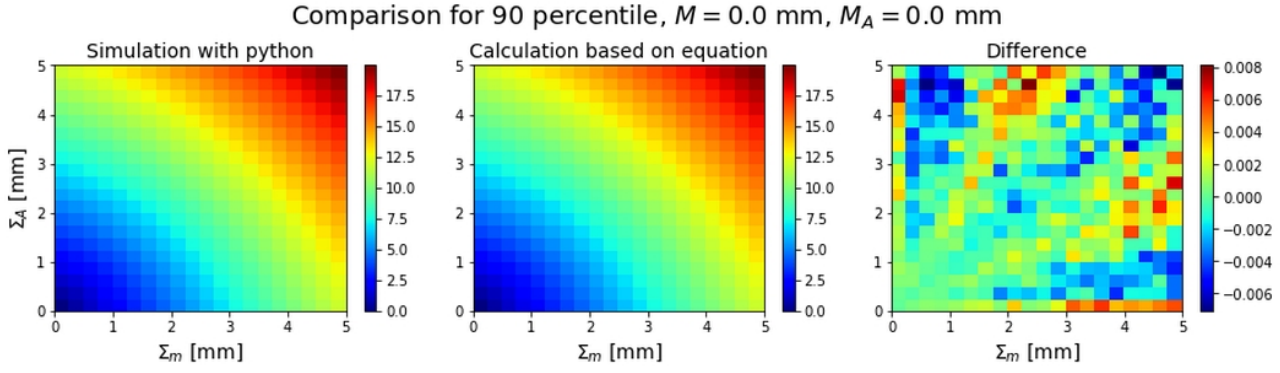


Figure 7.11: Comparison between margin calculation methods (Python code vs direct equation) for 90% of patients assumed to have all non-random error within the margin.

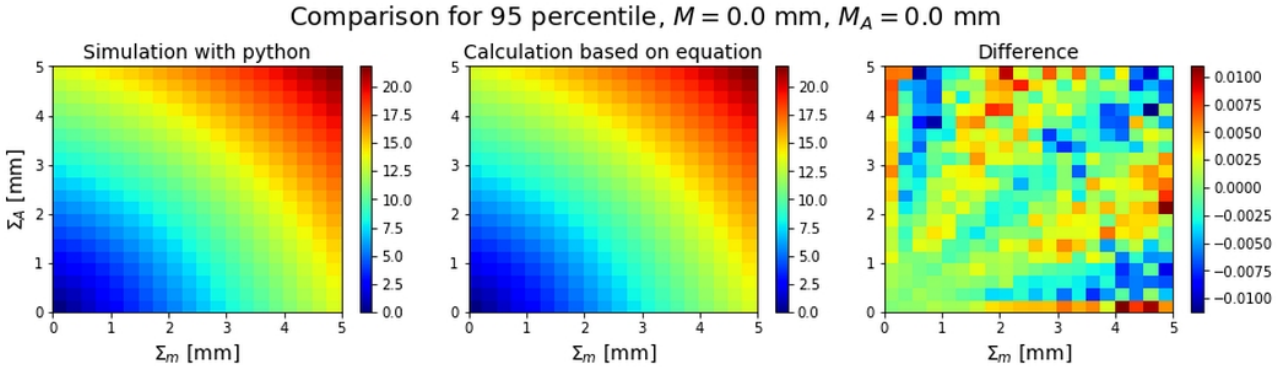


Figure 7.12: Comparison between margin calculation methods (Python code vs direct equation) for 95% of patients assumed to have all non-random error within the margin.

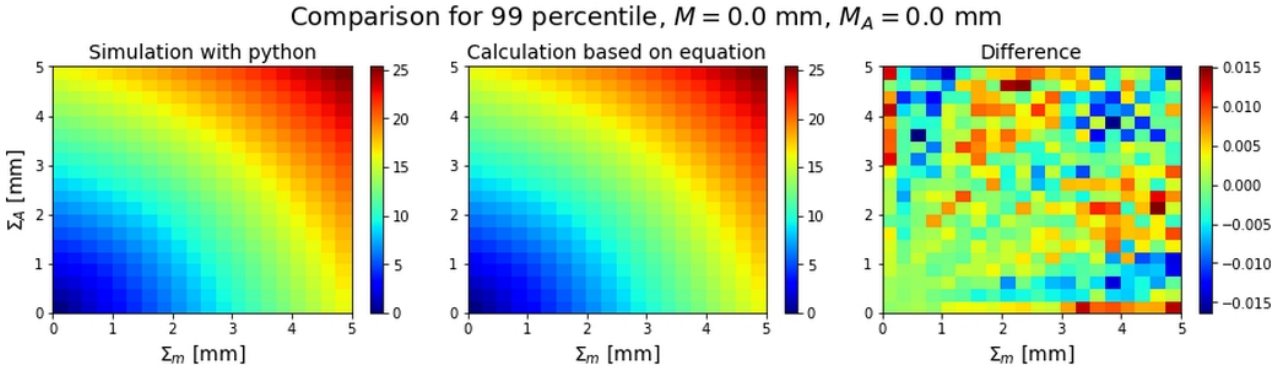


Figure 7.13: Comparison between margin calculation methods (Python code vs direct equation) for 99% of patients assumed to have all non-random error within the margin.

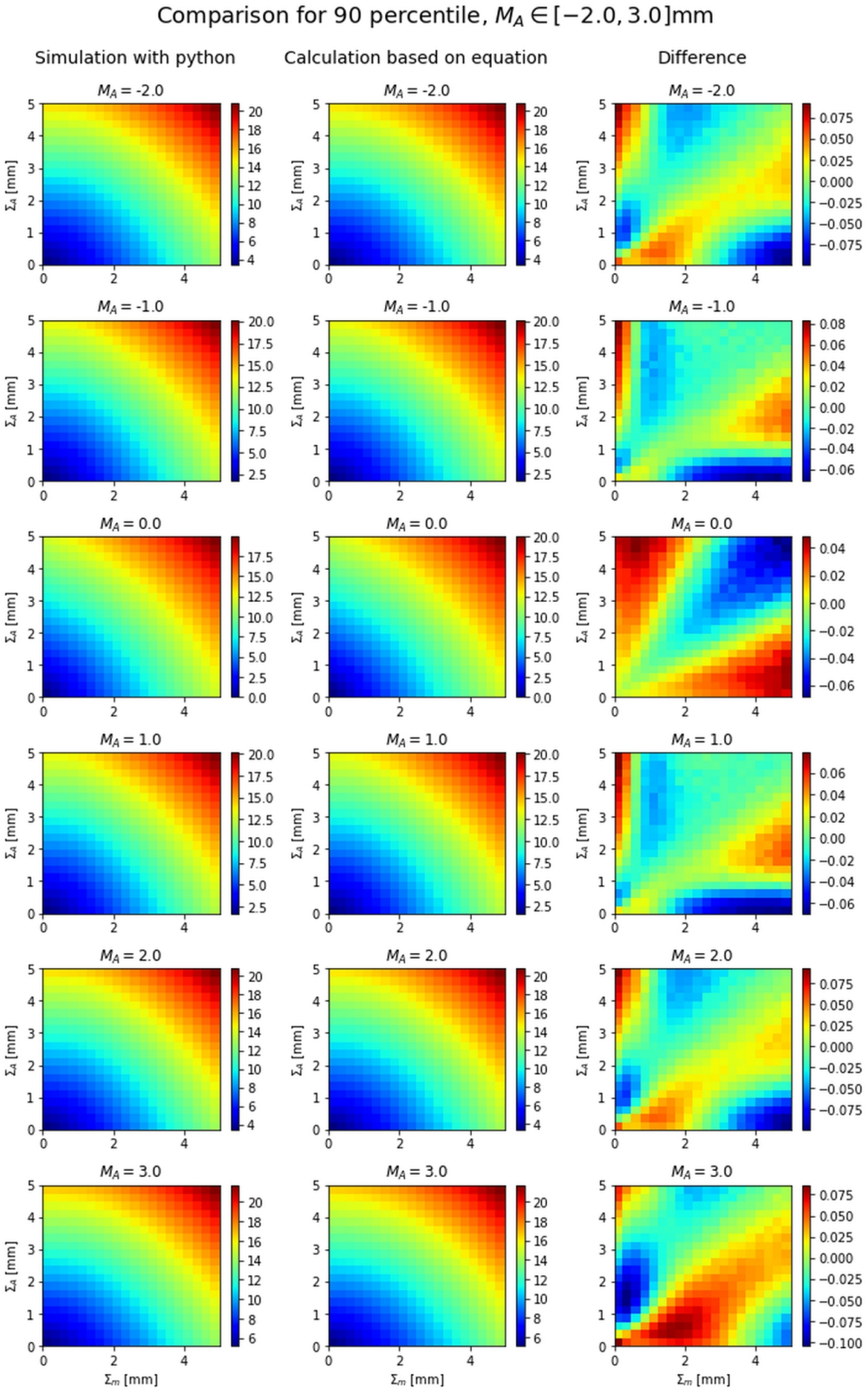


Figure 7.14: Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin.

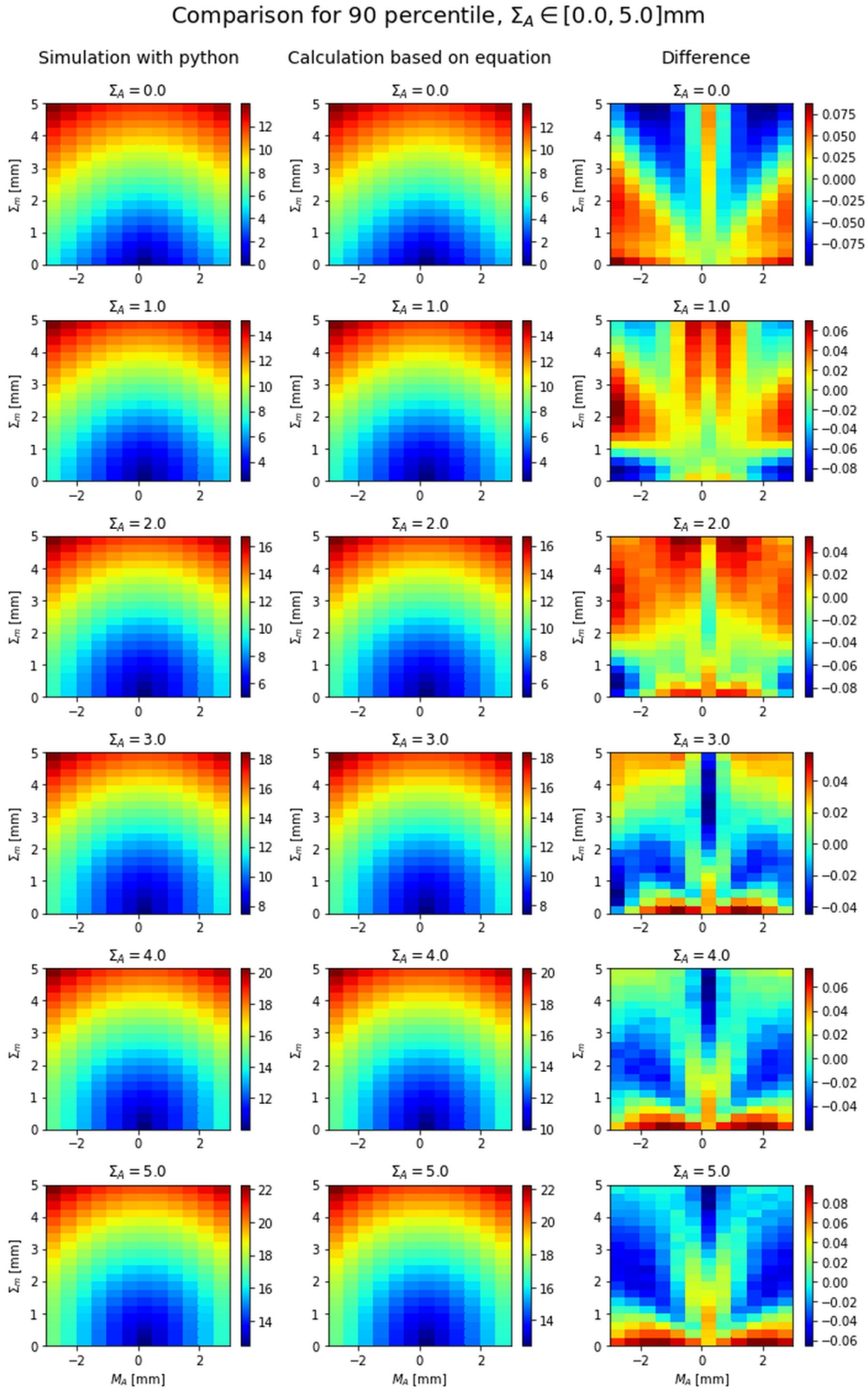


Figure 7.15: Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin.

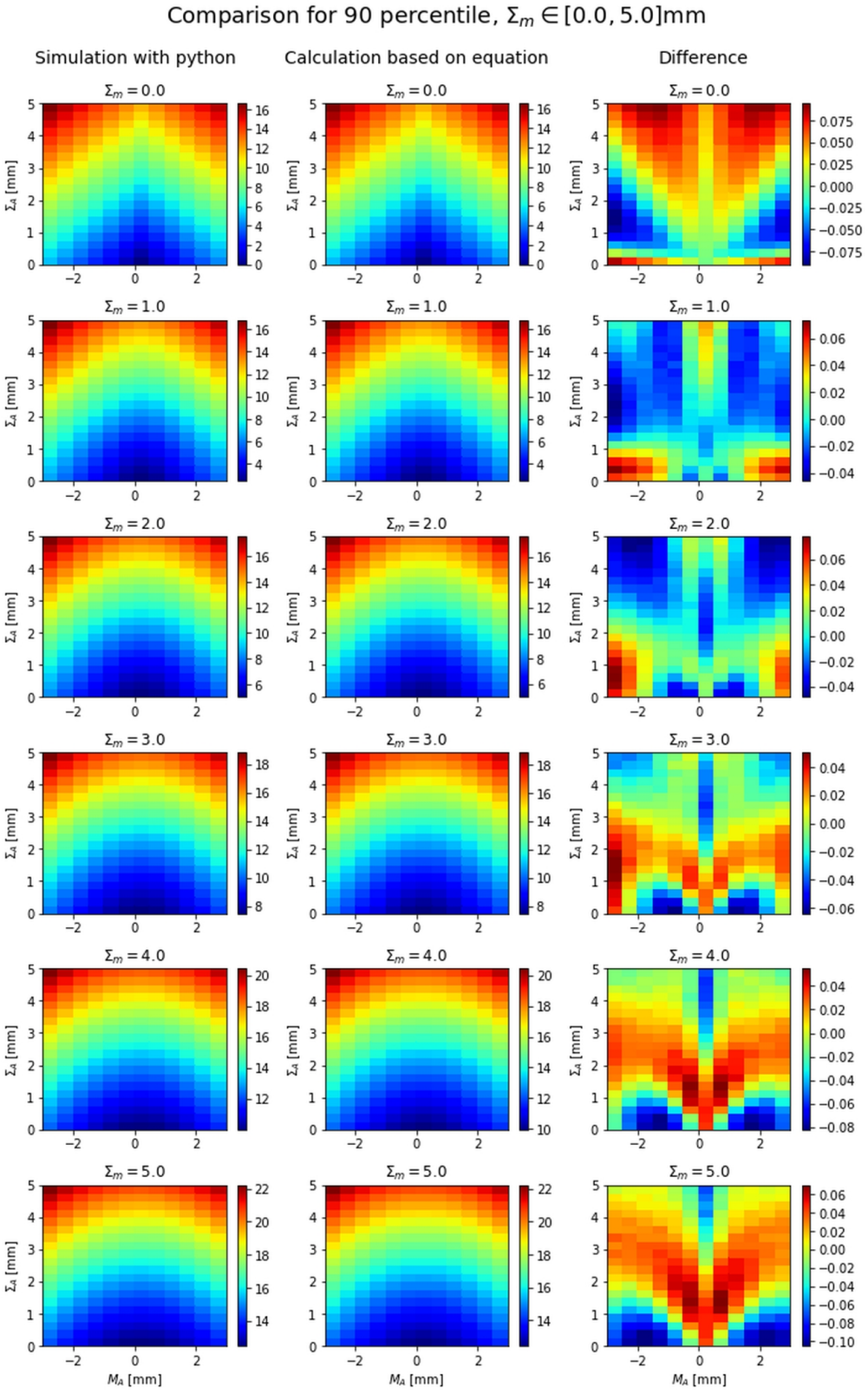


Figure 7.16: Comparison between margin calculation methods (Python code vs 4D direct equation) for 90% of patients assumed to have all non-random error within the margin. $\Sigma_m = \Sigma$

7.4 Discussion

To the best of my knowledge this is the first attempt to provide margin recipe that explicitly copes with inter-fraction time trends. In presented work I showed that van Herk's recipe in case of trends gives smaller margin which does not fulfill the assumption of 90% of patients inside the margin. Both recipes were tested for synthetic populations (isotropic and anisotropic) and for clinical population.

Trend-margin problem was considered in case of intra-fraction motion [75]. Authors proposed that Σ values achieved for localization accuracy, intra-fraction motion and respiratory motion shall be added in quadratic way. That was done with the assumption of on-line verification at the beginning of treatment (giving initial setup error of zero value). In worst case scenario intra-fraction motion would happen immediately after on-line verification.

In proposed approach we do not assume ideal patient position in first fraction of treatment. Therefore it can happen that error during time of treatment would become smaller or greater. Let $m_{p,i}(f)$ denote setup error in fraction f, of patient p in direction i. In case of time trends $m_{p,i}(1)$ can be used as a 'classic' systematic component. Distribution of $m_{p,i}(1)$ over the population of patients P would be described with the mean value (surrogate of GM for van Herk): $\overline{m_{p,i}(1)}$ and standard deviation: $\Sigma_{m_{p,i}(1)}$. The difference between first and last fraction can be expressed as: $\Delta m_{p,i}(F) = m_{p,i}(F) - m_{p,i}(1) = (F-1) \cdot a_{p,i}$.

Distribution of $\Delta m_{p,i}(F)$ can be also evaluated. Mean value of $\Delta m_{p,i}(F)$ would be: $\overline{\Delta m_{p,i}(F)}$ and its standard deviation: $\Sigma_{\Delta m_{p,i}(F)}$. The setup error in last fraction $m_{p,i}(F)$ can be expressed as $m_{p,i}(1)+\Delta m_{p,i}(F)$. Standard deviation of $m_{p,i}(F)$ can be calculated as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Although mathematically it is corrected we cannot use it to calculate the margin. It is because searched margin has to be large enough to have both $m_{p,i}(1)$ and $m_{p,i}(F)$ inside in 90% of patients. In Section 3.5 I showed that in some cases the $|m_{p,i}(F)|$ would be greater but in others $|m_{p,i}(1)|$ would be greater. That is why the greater value of the two: $m_{p,i}(1)$, $m_{p,i}(F)$) has to be searched. Fig. 7.17-7.19 show comparison between discussed methods. A drop in distributions of $max\{|m_p(0)|, |m_p(F)|\}$ can be observed for all principal directions. Distribution of $m_p(F)$ would be equal to $max\{|m_p(0)|, |m_p(F)|\}$ only if $|m_{p,i}(F)| \geq |m_{p,i}(1)|$, which is not always the case.

Second idea of calculating margins in terms of time trends would be to calculate middle position (see Eq. 7.4), as in proposed approach, and afterwards the extend from middle position to first/last fraction. That could be expressed as:
$\Delta_1 = \overline{m_{p,i}} - m_p(1) = (F-1) \cdot a/2$
$\Delta_F = m_p(F) - \overline{m_{p,i}} = (F-1) \cdot a/2$
As the Δ_1 and Δ_F are expressed in the same way, they are highly dependent and therefore cannot be added in quadratic way.

In 2015 Namysł-Kaletka et al. [67] investigated gastric cancer patients. They reported inter-fraction time trends up to 1.6mm during entire treatment. Still they used classical van Herk's recipe to derive margins.

Nowadays the on-line verification has a major role during radiotherapy. That become possible due to the technological progress. With on-line verification approach time trends can be resolved unless there are two target volumes which move independently. In such a case even if on-line verification assures correct position of one target the other one(s) can experience time trends. Such independent movement was reported for breast [29] and laryngeal irradiation [72]. Similar behavior can be expected for lung cancer [93].

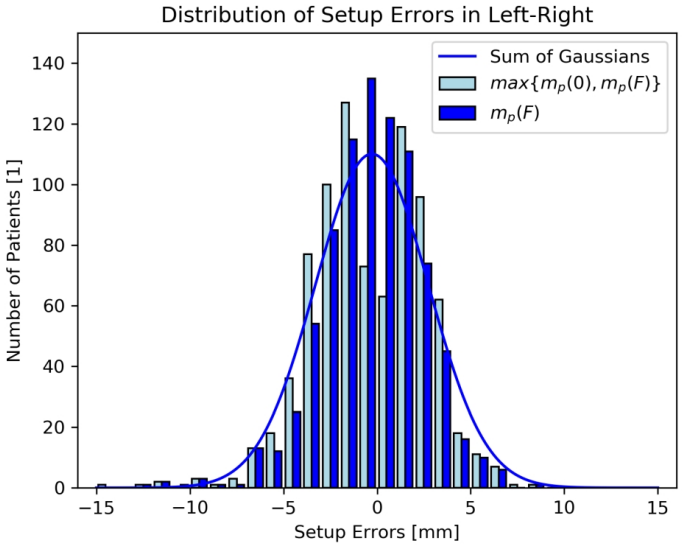


Figure 7.17: Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in left-right direction.

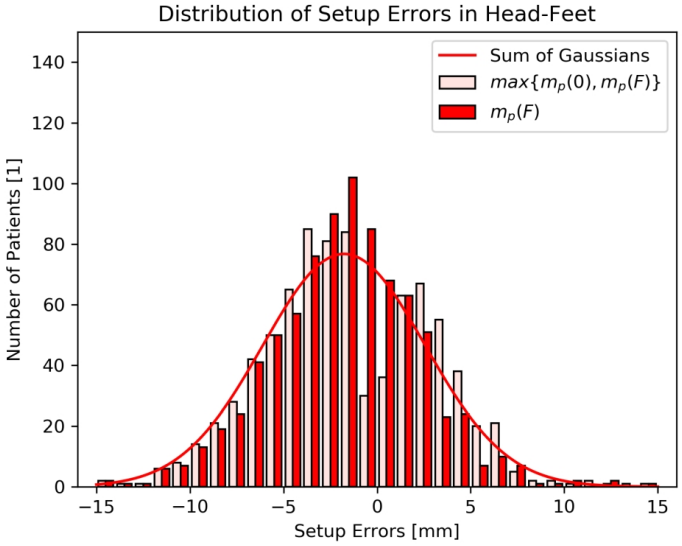


Figure 7.18: Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in head-feet direction.

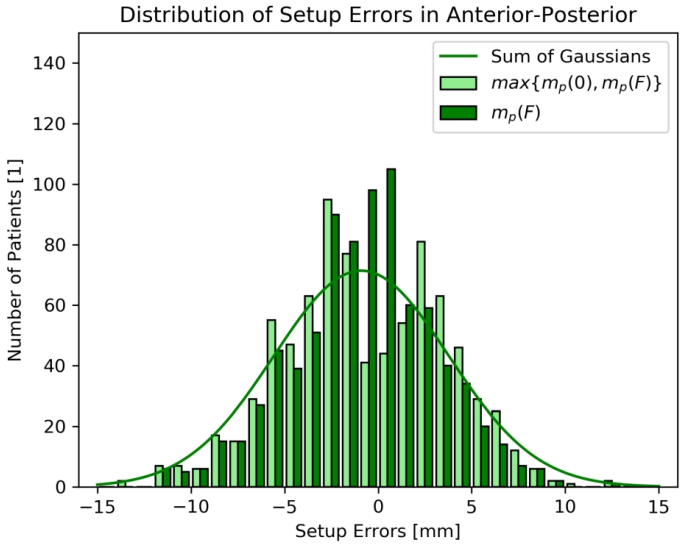


Figure 7.19: Comparison of discussed methods of calculating margin in case of time trends. Distribution of time trend setup error (without random component) in last fraction $m_p(F)$ is shown in blue. Second distribution presents $max\{|m_p(0)|, |m_p(F)|\}$ with the sign of $m_p(0)$ or $m_p(F)$. Gaussian plot was calculated with the SD given as $\sqrt{\Sigma^2_{m_{p,i}(1)} + \Sigma^2_{\Delta m_{p,i}(F)}}$. Presented data come from Erasmus MC database — in anterior-posterior direction.

Proposed recipe is based on folded normal distributions of middle setup error and trend slope. Although it is possible to give equation for time trend margin calculation (see Section 7.3.3) the form of these equation is sophisticated especially in case $M_A \neq 0$. That is why I propose to use look-up tables instead (Section 7.5).

7.5 Look-up tables

Tables 7.5-7.9 present margin values for different Σ_m and Σ_a. Simulations were done for 10^7 cases. Both mean middle position and mean slope were set to zero. Presented tables can be used in order to interpolate the margin for real patient population. First row of each presented table corresponds to van Herk recipe and is in agreement with it. It should be noticed that even in case of small Σ_a, which might be related to limited number of fractions and not to physiological time trends (see Section 3.6), the difference in margin size may be relatively big. For example in case of 90% requirement $\Sigma_a = 0.02$ mm/fraction leads to 0.5 mm change in margin size.

Tables 7.10-7.16 present values of the time trend margin for different M_A (see Eq. 7.8), Σ_A (see Eq. 7.9) and Σ_M with assumption that 90% of patients would have all non-random errors within the margin.

Σ_a \ Σ_m	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
0	0.0	1.1	2.2	3.2	4.3	5.4	6.5	7.5	8.6	9.7	10.8
0.01	0.4	1.3	2.4	3.5	4.5	5.6	6.7	7.8	8.9	9.9	11.0
0.02	0.8	1.7	2.7	3.7	4.8	5.9	6.9	8.0	9.1	10.2	11.2
0.03	1.2	2.0	3.0	4.0	5.1	6.1	7.2	8.3	9.3	10.4	11.5
0.04	1.6	2.4	3.3	4.3	5.4	6.4	7.5	8.5	9.6	10.7	11.7
0.05	2.0	2.8	3.7	4.6	5.7	6.7	7.7	8.8	9.9	10.9	12.0
0.06	2.5	3.2	4.0	5.0	6.0	7.0	8.0	9.1	10.1	11.2	12.3
0.07	2.9	3.6	4.4	5.3	6.3	7.3	8.3	9.4	10.4	11.5	12.5
0.08	3.3	4.0	4.8	5.7	6.6	7.6	8.6	9.7	10.7	11.8	12.8
0.09	3.7	4.4	5.1	6.0	7.0	7.9	8.9	10.0	11.0	12.0	13.1
0.1	4.1	4.8	5.5	6.4	7.3	8.3	9.3	10.3	11.3	12.3	13.4

Table 7.5: Margin values for requirement of 80% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 80% gives $2.16\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction.

Σ_a \ Σ_m	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
0	0.0	1.2	2.3	3.5	4.6	5.8	6.9	8.1	9.2	10.4	11.5
0.01	0.4	1.4	2.6	3.7	4.8	6.0	7.2	8.3	9.5	10.6	11.8
0.02	0.9	1.7	2.8	4.0	5.1	6.2	7.4	8.5	9.7	10.8	12.0
0.03	1.3	2.1	3.1	4.2	5.4	6.5	7.7	8.8	9.9	11.1	12.2
0.04	1.8	2.5	3.5	4.6	5.7	6.8	7.9	9.1	10.2	11.4	12.5
0.05	2.2	2.9	3.8	4.9	6.0	7.1	8.2	9.3	10.5	11.6	12.8
0.06	2.6	3.3	4.2	5.2	6.3	7.4	8.5	9.6	10.8	11.9	13.0
0.07	3.1	3.8	4.6	5.6	6.6	7.7	8.8	9.9	11.0	12.2	13.3
0.08	3.5	4.2	5.0	6.0	7.0	8.0	9.1	10.2	11.3	12.5	13.6
0.09	3.9	4.6	5.4	6.3	7.3	8.4	9.4	10.5	11.6	12.8	13.9
0.1	4.4	5.0	5.8	6.7	7.7	8.7	9.8	10.9	11.9	13.1	14.2

Table 7.6: Margin values for requirement of 85% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 85% gives $2.31\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction.

Σ_a \ Σ_m	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
0	0.0	1.3	2.5	3.8	5.0	6.3	7.5	8.8	10.0	11.3	12.5
0.01	0.5	1.5	2.7	4.0	5.2	6.5	7.7	9.0	10.2	11.5	12.7
0.02	1.0	1.9	3.0	4.3	5.5	6.7	8.0	9.2	10.5	11.7	13.0
0.03	1.4	2.2	3.4	4.6	5.8	7.0	8.2	9.5	10.7	12.0	13.2
0.04	1.9	2.7	3.7	4.9	6.1	7.3	8.5	9.8	11.0	12.2	13.5
0.05	2.4	3.1	4.1	5.2	6.4	7.6	8.8	10.0	11.3	12.5	13.7
0.06	2.9	3.6	4.5	5.6	6.7	7.9	9.1	10.3	11.5	12.8	14.0
0.07	3.3	4.0	4.9	5.9	7.1	8.2	9.4	10.6	11.8	13.1	14.3
0.08	3.8	4.5	5.3	6.3	7.4	8.6	9.7	10.9	12.1	13.4	14.6
0.09	4.3	5.0	5.8	6.7	7.8	8.9	10.1	11.3	12.4	13.7	14.9
0.1	4.8	5.4	6.2	7.2	8.2	9.3	10.4	11.6	12.8	14.0	15.2

Table 7.7: Margin values for requirement of 90% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 90% gives $2.50\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction.

Σ_a \ Σ_m	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
0	0.0	1.4	2.8	4.2	5.6	7.0	8.4	9.8	11.2	12.6	14.0
0.01	0.5	1.7	3.0	4.4	5.8	7.2	8.6	10.0	11.4	12.8	14.2
0.02	1.1	2.0	3.3	4.7	6.1	7.5	8.9	10.3	11.7	13.1	14.4
0.03	1.6	2.5	3.7	5.0	6.4	7.7	9.1	10.5	11.9	13.3	14.7
0.04	2.1	2.9	4.1	5.3	6.7	8.0	9.4	10.8	12.2	13.6	15.0
0.05	2.7	3.4	4.5	5.7	7.0	8.3	9.7	11.1	12.5	13.8	15.2
0.06	3.2	3.9	4.9	6.1	7.4	8.7	10.0	11.4	12.7	14.1	15.5
0.07	3.7	4.4	5.4	6.5	7.7	9.0	10.3	11.7	13.0	14.4	15.8
0.08	4.3	4.9	5.9	6.9	8.1	9.4	10.7	12.0	13.4	14.7	16.1
0.09	4.8	5.5	6.3	7.4	8.5	9.8	11.0	12.3	13.7	15.0	16.4
0.1	5.3	6.0	6.8	7.8	8.9	10.1	11.4	12.7	14.0	15.4	16.7

Table 7.8: Margin values for requirement of 95% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 95% gives $2.79\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction.

Σ_a \ Σ_m	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
0	0.0	1.7	3.4	5.1	6.7	8.4	10.1	11.8	13.5	15.2	16.8
0.01	0.6	2.0	3.6	5.3	7.0	8.7	10.3	12.0	13.7	15.4	17.1
0.02	1.3	2.4	3.9	5.6	7.2	8.9	10.6	12.3	13.9	15.6	17.3
0.03	1.9	2.9	4.3	5.9	7.5	9.2	10.9	12.5	14.2	15.9	17.6
0.04	2.6	3.4	4.7	6.3	7.9	9.5	11.2	12.8	14.5	16.2	17.8
0.05	3.2	4.0	5.2	6.7	8.2	9.8	11.5	13.1	14.8	16.4	18.1
0.06	3.8	4.6	5.7	7.1	8.6	10.2	11.8	13.4	15.1	16.7	18.4
0.07	4.5	5.2	6.3	7.6	9.0	10.6	12.1	13.8	15.4	17.0	18.7
0.08	5.1	5.8	6.8	8.1	9.5	11.0	12.5	14.1	15.7	17.4	19.0
0.09	5.8	6.5	7.4	8.6	9.9	11.4	12.9	14.5	16.1	17.7	19.3
0.1	6.4	7.1	8.0	9.1	10.4	11.8	13.3	14.9	16.4	18.0	19.7

Table 7.9: Margin values for requirement of 99% of patient population to have CTV within the prescribed dose. Margins were calculated for different SD of middle position and slope with M and M_a set to zero. First row of the presented table corresponds to van Herk margin recipe, which in case of 99% gives $3.36\Sigma_m$. The Σ_m values are given in mm and Σ_a values are given in mm/fraction.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	0.0	0.6	1.3	1.9	2.5	3.1	3.8	4.4	5.0	5.6	6.2	6.9	7.5	8.1	8.8	9.4	10.0	10.6	11.3	11.9	12.5
0.25	0.6	1.1	1.6	2.2	2.8	3.5	4.1	4.7	5.3	5.9	6.6	7.2	7.8	8.4	9.1	9.7	10.3	10.9	11.6	12.2	12.8
0.5	1.3	1.6	2.1	2.7	3.2	3.8	4.4	5.1	5.7	6.3	6.9	7.5	8.1	8.8	9.4	10.0	10.6	11.3	11.9	12.5	13.1
0.75	1.9	2.2	2.7	3.2	3.7	4.3	4.9	5.4	6.1	6.7	7.3	7.9	8.5	9.1	9.7	10.4	11.0	11.6	12.2	12.8	13.5
1	2.5	2.8	3.2	3.7	4.2	4.7	5.3	5.9	6.5	7.1	7.7	8.3	8.9	9.5	10.1	10.7	11.3	12.0	12.6	13.2	13.8
1.25	3.1	3.5	3.8	4.3	4.7	5.3	5.8	6.3	6.9	7.5	8.1	8.7	9.3	9.9	10.5	11.1	11.7	12.3	12.9	13.6	14.2
1.5	3.8	4.1	4.4	4.9	5.3	5.8	6.3	6.8	7.4	8.0	8.5	9.1	9.7	10.3	10.9	11.5	12.1	12.7	13.3	13.9	14.5
1.75	4.4	4.7	5.1	5.4	5.9	6.3	6.8	7.4	7.9	8.4	9.0	9.6	10.1	10.7	11.3	11.9	12.5	13.1	13.7	14.3	14.9
2	5.0	5.3	5.7	6.1	6.5	6.9	7.4	7.9	8.4	8.9	9.5	10.0	10.6	11.2	11.8	12.3	12.9	13.5	14.1	14.7	15.3
2.25	5.6	5.9	6.3	6.7	7.1	7.5	8.0	8.4	8.9	9.5	10.0	10.5	11.1	11.6	12.2	12.8	13.4	14.0	14.6	15.1	15.7
2.5	6.3	6.6	6.9	7.3	7.7	8.1	8.5	9.0	9.5	10.0	10.5	11.0	11.6	12.1	12.7	13.3	13.8	14.4	15.0	15.6	16.2
2.75	6.9	7.2	7.5	7.9	8.3	8.7	9.1	9.6	10.0	10.5	11.0	11.6	12.1	12.6	13.2	13.7	14.3	14.9	15.4	16.0	16.6
3	7.5	7.8	8.1	8.5	8.9	9.3	9.7	10.1	10.6	11.1	11.6	12.1	12.6	13.1	13.7	14.2	14.8	15.3	15.9	16.5	17.1
3.25	8.1	8.4	8.8	9.1	9.5	9.9	10.3	10.7	11.2	11.6	12.1	12.6	13.1	13.7	14.2	14.7	15.3	15.8	16.4	16.9	17.5
3.5	8.7	9.1	9.4	9.7	10.1	10.5	10.9	11.3	11.8	12.2	12.7	13.2	13.7	14.2	14.7	15.2	15.8	16.3	16.9	17.4	18.0
3.75	9.4	9.7	10.0	10.4	10.7	11.1	11.5	11.9	12.3	12.8	13.3	13.7	14.2	14.7	15.2	15.8	16.3	16.8	17.4	17.9	18.5
4	10.0	10.3	10.6	11.0	11.3	11.7	12.1	12.5	12.9	13.4	13.8	14.3	14.8	15.3	15.8	16.3	16.8	17.3	17.9	18.4	19.0
4.25	10.6	10.9	11.3	11.6	11.9	12.3	12.7	13.1	13.5	14.0	14.4	14.9	15.3	15.8	16.3	16.8	17.3	17.9	18.4	18.9	19.5
4.5	11.3	11.6	11.9	12.2	12.6	12.9	13.3	13.7	14.1	14.6	15.0	15.4	15.9	16.4	16.9	17.4	17.9	18.4	18.9	19.4	20.0
4.75	11.9	12.2	12.5	12.8	13.2	13.6	13.9	14.3	14.7	15.1	15.6	16.0	16.5	16.9	17.4	17.9	18.4	18.9	19.4	20.0	20.5
5	12.5	12.8	13.1	13.5	13.8	14.2	14.5	14.9	15.3	15.7	16.2	16.6	17.1	17.5	18.0	18.5	19.0	19.5	20.0	20.5	21.0

Table 7.10: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = 0$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	0.9	1.4	2.0	2.7	3.3	3.9	4.5	5.1	5.8	6.4	7.0	7.6	8.3	8.9	9.5	10.1	10.8	11.4	12.0	12.6	13.3
0.25	1.2	1.7	2.2	2.8	3.4	4.0	4.6	5.2	5.8	6.4	7.0	7.7	8.3	8.9	9.5	10.2	10.8	11.4	12.0	12.7	13.3
0.5	1.7	2.1	2.5	3.0	3.6	4.2	4.8	5.4	6.0	6.6	7.2	7.8	8.5	9.1	9.7	10.3	10.9	11.6	12.2	12.8	13.4
0.75	2.2	2.6	3.0	3.5	4.0	4.5	5.1	5.7	6.3	6.9	7.5	8.1	8.7	9.4	10.0	10.6	11.2	11.8	12.4	13.1	13.7
1	2.8	3.1	3.5	4.0	4.4	5.0	5.5	6.1	6.7	7.3	7.9	8.5	9.1	9.7	10.3	10.9	11.5	12.1	12.7	13.4	14.0
1.25	3.4	3.7	4.1	4.5	5.0	5.5	6.0	6.5	7.1	7.7	8.3	8.8	9.4	10.0	10.6	11.3	11.9	12.5	13.1	13.7	14.3
1.5	3.9	4.3	4.6	5.0	5.5	6.0	6.5	7.0	7.5	8.1	8.7	9.3	9.8	10.4	11.0	11.6	12.2	12.8	13.4	14.1	14.7
1.75	4.5	4.9	5.2	5.6	6.0	6.5	7.0	7.5	8.0	8.6	9.1	9.7	10.3	10.9	11.4	12.0	12.6	13.2	13.8	14.4	15.0
2	5.2	5.5	5.8	6.2	6.6	7.1	7.5	8.0	8.5	9.1	9.6	10.2	10.7	11.3	11.9	12.5	13.0	13.6	14.2	14.8	15.4
2.25	5.8	6.1	6.4	6.8	7.2	7.6	8.1	8.6	9.1	9.6	10.1	10.6	11.2	11.8	12.3	12.9	13.5	14.1	14.6	15.2	15.8
2.5	6.4	6.7	7.0	7.4	7.8	8.2	8.6	9.1	9.6	10.1	10.6	11.1	11.7	12.2	12.8	13.4	13.9	14.5	15.1	15.7	16.3
2.75	7.0	7.3	7.6	8.0	8.4	8.8	9.2	9.7	10.1	10.6	11.1	11.6	12.2	12.7	13.3	13.8	14.4	15.0	15.5	16.1	16.7
3	7.6	7.9	8.2	8.6	9.0	9.4	9.8	10.2	10.7	11.2	11.7	12.2	12.7	13.2	13.8	14.3	14.9	15.4	16.0	16.6	17.1
3.25	8.2	8.5	8.9	9.2	9.6	10.0	10.4	10.8	11.3	11.7	12.2	12.7	13.2	13.7	14.3	14.8	15.3	15.9	16.5	17.0	17.6
3.5	8.8	9.1	9.5	9.8	10.2	10.6	11.0	11.4	11.8	12.3	12.8	13.2	13.7	14.3	14.8	15.3	15.8	16.4	16.9	17.5	18.1
3.75	9.5	9.8	10.1	10.4	10.8	11.2	11.6	12.0	12.4	12.9	13.3	13.8	14.3	14.8	15.3	15.8	16.4	16.9	17.4	18.0	18.5
4	10.1	10.4	10.7	11.1	11.4	11.8	12.2	12.6	13.0	13.4	13.9	14.4	14.8	15.3	15.8	16.4	16.9	17.4	17.9	18.5	19.0
4.25	10.7	11.0	11.3	11.7	12.0	12.4	12.8	13.2	13.6	14.0	14.5	14.9	15.4	15.9	16.4	16.9	17.4	17.9	18.4	19.0	19.5
4.5	11.3	11.6	11.9	12.3	12.6	13.0	13.4	13.8	14.2	14.6	15.1	15.5	16.0	16.4	16.9	17.4	17.9	18.4	19.0	19.5	20.0
4.75	11.9	12.2	12.6	12.9	13.3	13.6	14.0	14.4	14.8	15.2	15.6	16.1	16.5	17.0	17.5	18.0	18.5	19.0	19.5	20.0	20.5
5	12.6	12.9	13.2	13.5	13.9	14.2	14.6	15.0	15.4	15.8	16.2	16.7	17.1	17.6	18.0	18.5	19.0	19.5	20.0	20.5	21.1

Table 7.11: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 0.5$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	1.7	2.3	2.9	3.5	4.1	4.7	5.3	5.9	6.6	7.2	7.8	8.4	9.0	9.7	10.3	10.9	11.5	12.2	12.8	13.4	14.0
0.25	2.1	2.5	3.0	3.6	4.2	4.8	5.4	6.0	6.6	7.2	7.8	8.5	9.1	9.7	10.3	10.9	11.6	12.2	12.8	13.4	14.0
0.5	2.5	2.9	3.3	3.8	4.4	4.9	5.5	6.1	6.7	7.3	7.9	8.5	9.2	9.8	10.4	11.0	11.6	12.2	12.9	13.5	14.1
0.75	2.9	3.3	3.7	4.2	4.7	5.2	5.8	6.3	6.9	7.5	8.1	8.7	9.3	9.9	10.5	11.2	11.8	12.4	13.0	13.6	14.2
1	3.4	3.8	4.1	4.6	5.1	5.6	6.1	6.6	7.2	7.8	8.4	9.0	9.6	10.2	10.8	11.4	12.0	12.6	13.2	13.8	14.4
1.25	3.9	4.3	4.6	5.0	5.5	6.0	6.5	7.0	7.6	8.1	8.7	9.3	9.9	10.5	11.1	11.7	12.3	12.9	13.5	14.1	14.7
1.5	4.5	4.8	5.1	5.5	6.0	6.4	6.9	7.4	8.0	8.5	9.1	9.7	10.2	10.8	11.4	12.0	12.6	13.2	13.8	14.4	15.0
1.75	5.0	5.3	5.7	6.1	6.5	6.9	7.4	7.9	8.4	9.0	9.5	10.1	10.6	11.2	11.8	12.4	13.0	13.6	14.2	14.8	15.4
2	5.6	5.9	6.2	6.6	7.0	7.5	7.9	8.4	8.9	9.4	10.0	10.5	11.1	11.6	12.2	12.8	13.4	13.9	14.5	15.1	15.7
2.25	6.1	6.5	6.8	7.2	7.6	8.0	8.4	8.9	9.4	9.9	10.4	11.0	11.5	12.1	12.6	13.2	13.8	14.3	14.9	15.5	16.1
2.5	6.7	7.0	7.4	7.7	8.1	8.5	9.0	9.4	9.9	10.4	10.9	11.4	12.0	12.5	13.1	13.6	14.2	14.8	15.3	15.9	16.5
2.75	7.3	7.6	8.0	8.3	8.7	9.1	9.5	10.0	10.4	10.9	11.4	11.9	12.4	13.0	13.5	14.1	14.6	15.2	15.8	16.3	16.9
3	7.9	8.2	8.5	8.9	9.3	9.7	10.1	10.5	11.0	11.4	11.9	12.4	12.9	13.5	14.0	14.5	15.1	15.7	16.2	16.8	17.4
3.25	8.5	8.8	9.1	9.5	9.9	10.2	10.7	11.1	11.5	12.0	12.5	12.9	13.5	14.0	14.5	15.0	15.6	16.1	16.7	17.2	17.8
3.5	9.1	9.4	9.7	10.1	10.4	10.8	11.2	11.7	12.1	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.1	16.6	17.1	17.7	18.3
3.75	9.7	10.0	10.3	10.7	11.0	11.4	11.8	12.2	12.7	13.1	13.6	14.0	14.5	15.0	15.5	16.0	16.6	17.1	17.6	18.2	18.7
4	10.3	10.6	10.9	11.3	11.6	12.0	12.4	12.8	13.2	13.7	14.1	14.6	15.1	15.5	16.0	16.5	17.1	17.6	18.1	18.7	19.2
4.25	10.9	11.2	11.5	11.9	12.2	12.6	13.0	13.4	13.8	14.2	14.7	15.1	15.6	16.1	16.6	17.1	17.6	18.1	18.6	19.2	19.7
4.5	11.5	11.8	12.2	12.5	12.8	13.2	13.6	14.0	14.4	14.8	15.3	15.7	16.2	16.6	17.1	17.6	18.1	18.6	19.1	19.7	20.2
4.75	12.1	12.4	12.8	13.1	13.4	13.8	14.2	14.6	15.0	15.4	15.8	16.3	16.7	17.2	17.7	18.1	18.6	19.1	19.7	20.2	20.7
5	12.7	13.1	13.4	13.7	14.1	14.4	14.8	15.2	15.6	16.0	16.4	16.8	17.3	17.7	18.2	18.7	19.2	19.7	20.2	20.7	21.2

Table 7.12: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 1.0$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	2.6	3.2	3.7	4.3	4.9	5.5	6.1	6.8	7.4	8.0	8.6	9.2	9.8	10.5	11.1	11.7	12.3	12.9	13.6	14.2	14.8
0.25	2.9	3.3	3.9	4.4	5.0	5.6	6.2	6.8	7.4	8.0	8.6	9.3	9.9	10.5	11.1	11.7	12.3	13.0	13.6	14.2	14.8
0.5	3.3	3.7	4.1	4.6	5.2	5.7	6.3	6.9	7.5	8.1	8.7	9.3	9.9	10.6	11.2	11.8	12.4	13.0	13.6	14.3	14.9
0.75	3.7	4.1	4.5	5.0	5.5	6.0	6.5	7.1	7.7	8.3	8.9	9.5	10.1	10.7	11.3	11.9	12.5	13.1	13.7	14.3	15.0
1	4.2	4.5	4.9	5.3	5.8	6.3	6.8	7.4	7.9	8.5	9.1	9.7	10.3	10.9	11.5	12.1	12.7	13.3	13.9	14.5	15.1
1.25	4.7	5.0	5.4	5.8	6.2	6.7	7.2	7.7	8.2	8.8	9.4	9.9	10.5	11.1	11.7	12.3	12.9	13.5	14.1	14.7	15.3
1.5	5.1	5.5	5.8	6.2	6.6	7.1	7.6	8.1	8.6	9.1	9.7	10.3	10.8	11.4	12.0	12.6	13.2	13.8	14.4	15.0	15.6
1.75	5.6	6.0	6.3	6.7	7.1	7.5	8.0	8.5	9.0	9.5	10.1	10.6	11.2	11.7	12.3	12.9	13.5	14.1	14.7	15.3	15.9
2	6.2	6.5	6.8	7.2	7.6	8.0	8.5	8.9	9.4	10.0	10.5	11.0	11.6	12.1	12.7	13.3	13.8	14.4	15.0	15.6	16.2
2.25	6.7	7.0	7.4	7.7	8.1	8.5	9.0	9.4	9.9	10.4	10.9	11.4	12.0	12.5	13.1	13.6	14.2	14.8	15.4	15.9	16.5
2.5	7.2	7.6	7.9	8.3	8.6	9.0	9.5	9.9	10.4	10.9	11.4	11.9	12.4	12.9	13.5	14.0	14.6	15.2	15.7	16.3	16.9
2.75	7.8	8.1	8.4	8.8	9.2	9.6	10.0	10.4	10.9	11.4	11.8	12.4	12.9	13.4	13.9	14.5	15.0	15.6	16.2	16.7	17.3
3	8.4	8.7	9.0	9.3	9.7	10.1	10.5	11.0	11.4	11.9	12.3	12.8	13.3	13.9	14.4	14.9	15.5	16.0	16.6	17.1	17.7
3.25	8.9	9.2	9.6	9.9	10.3	10.7	11.1	11.5	11.9	12.4	12.9	13.3	13.8	14.3	14.9	15.4	15.9	16.5	17.0	17.6	18.1
3.5	9.5	9.8	10.1	10.5	10.8	11.2	11.6	12.0	12.5	12.9	13.4	13.9	14.3	14.8	15.3	15.9	16.4	16.9	17.5	18.0	18.6
3.75	10.1	10.4	10.7	11.1	11.4	11.8	12.2	12.6	13.0	13.5	13.9	14.4	14.9	15.3	15.8	16.4	16.9	17.4	17.9	18.5	19.0
4	10.7	11.0	11.3	11.6	12.0	12.4	12.8	13.2	13.6	14.0	14.5	14.9	15.4	15.9	16.4	16.9	17.4	17.9	18.4	19.0	19.5
4.25	11.3	11.6	11.9	12.2	12.6	12.9	13.3	13.7	14.1	14.6	15.0	15.5	15.9	16.4	16.9	17.4	17.9	18.4	18.9	19.4	20.0
4.5	11.8	12.2	12.5	12.8	13.2	13.5	13.9	14.3	14.7	15.1	15.6	16.0	16.5	16.9	17.4	17.9	18.4	18.9	19.4	19.9	20.5
4.75	12.4	12.8	13.1	13.4	13.8	14.1	14.5	14.9	15.3	15.7	16.1	16.6	17.0	17.5	17.9	18.4	18.9	19.4	19.9	20.4	21.0
5	13.0	13.4	13.7	14.0	14.4	14.7	15.1	15.5	15.9	16.3	16.7	17.1	17.6	18.0	18.5	19.0	19.4	19.9	20.4	21.0	21.5

Table 7.13: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 1.5$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	3.5	4.0	4.6	5.2	5.8	6.4	7.0	7.6	8.2	8.8	9.4	10.0	10.7	11.3	11.9	12.5	13.1	13.7	14.4	15.0	15.6
0.25	3.8	4.2	4.7	5.3	5.8	6.4	7.0	7.6	8.2	8.8	9.4	10.1	10.7	11.3	11.9	12.5	13.1	13.8	14.4	15.0	15.6
0.5	4.2	4.5	5.0	5.5	6.0	6.6	7.2	7.7	8.3	8.9	9.5	10.1	10.7	11.4	12.0	12.6	13.2	13.8	14.4	15.0	15.7
0.75	4.6	4.9	5.3	5.8	6.3	6.8	7.4	7.9	8.5	9.1	9.7	10.3	10.9	11.5	12.1	12.7	13.3	13.9	14.5	15.1	15.7
1	5.0	5.3	5.7	6.2	6.6	7.1	7.6	8.2	8.7	9.3	9.9	10.4	11.0	11.6	12.2	12.8	13.4	14.0	14.6	15.2	15.9
1.25	5.4	5.8	6.1	6.6	7.0	7.5	8.0	8.5	9.0	9.6	10.1	10.7	11.3	11.8	12.4	13.0	13.6	14.2	14.8	15.4	16.0
1.5	5.9	6.2	6.6	7.0	7.4	7.8	8.3	8.8	9.3	9.9	10.4	11.0	11.5	12.1	12.7	13.3	13.8	14.4	15.0	15.6	16.2
1.75	6.4	6.7	7.0	7.4	7.8	8.3	8.7	9.2	9.7	10.2	10.7	11.3	11.8	12.4	13.0	13.5	14.1	14.7	15.3	15.9	16.5
2	6.9	7.2	7.5	7.9	8.3	8.7	9.2	9.6	10.1	10.6	11.1	11.6	12.2	12.7	13.3	13.9	14.4	15.0	15.6	16.2	16.8
2.25	7.4	7.7	8.0	8.4	8.8	9.2	9.6	10.1	10.5	11.0	11.5	12.0	12.6	13.1	13.7	14.2	14.8	15.3	15.9	16.5	17.1
2.5	7.9	8.2	8.5	8.9	9.3	9.7	10.1	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.6	15.2	15.7	16.3	16.8	17.4
2.75	8.4	8.7	9.0	9.4	9.8	10.2	10.6	11.0	11.5	11.9	12.4	12.9	13.4	13.9	14.5	15.0	15.5	16.1	16.7	17.2	17.8
3	8.9	9.2	9.6	9.9	10.3	10.7	11.1	11.5	12.0	12.4	12.9	13.4	13.9	14.4	14.9	15.4	16.0	16.5	17.1	17.6	18.2
3.25	9.5	9.8	10.1	10.5	10.8	11.2	11.6	12.0	12.5	12.9	13.4	13.8	14.3	14.8	15.3	15.9	16.4	16.9	17.5	18.0	18.6
3.5	10.0	10.3	10.7	11.0	11.4	11.7	12.1	12.5	13.0	13.4	13.9	14.3	14.8	15.3	15.8	16.3	16.8	17.4	17.9	18.5	19.0
3.75	10.6	10.9	11.2	11.6	11.9	12.3	12.7	13.1	13.5	13.9	14.4	14.8	15.3	15.8	16.3	16.8	17.3	17.8	18.4	18.9	19.4
4	11.1	11.5	11.8	12.1	12.5	12.8	13.2	13.6	14.0	14.5	14.9	15.4	15.8	16.3	16.8	17.3	17.8	18.3	18.8	19.4	19.9
4.25	11.7	12.0	12.3	12.7	13.0	13.4	13.8	14.2	14.6	15.0	15.4	15.9	16.3	16.8	17.3	17.8	18.3	18.8	19.3	19.8	20.4
4.5	12.3	12.6	12.9	13.2	13.6	14.0	14.3	14.7	15.1	15.5	16.0	16.4	16.9	17.3	17.8	18.3	18.8	19.3	19.8	20.3	20.8
4.75	12.9	13.2	13.5	13.8	14.2	14.5	14.9	15.3	15.7	16.1	16.5	17.0	17.4	17.9	18.3	18.8	19.3	19.8	20.3	20.8	21.3
5	13.4	13.7	14.1	14.4	14.7	15.1	15.5	15.9	16.2	16.7	17.1	17.5	17.9	18.4	18.9	19.3	19.8	20.3	20.8	21.3	21.8

Table 7.14: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 2.0$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	4.3	4.9	5.5	6.0	6.6	7.2	7.8	8.4	9.0	9.6	10.2	10.9	11.5	12.1	12.7	13.3	13.9	14.6	15.2	15.8	16.4
0.25	4.7	5.1	5.6	6.1	6.7	7.3	7.9	8.5	9.1	9.7	10.3	10.9	11.5	12.1	12.7	13.3	14.0	14.6	15.2	15.8	16.4
0.5	5.0	5.4	5.8	6.3	6.9	7.4	8.0	8.6	9.2	9.8	10.4	11.0	11.6	12.2	12.8	13.4	14.0	14.6	15.2	15.8	16.5
0.75	5.4	5.8	6.2	6.6	7.1	7.7	8.2	8.8	9.3	9.9	10.5	11.1	11.7	12.3	12.9	13.5	14.1	14.7	15.3	15.9	16.5
1	5.8	6.2	6.6	7.0	7.4	7.9	8.5	9.0	9.5	10.1	10.7	11.2	11.8	12.4	13.0	13.6	14.2	14.8	15.4	16.0	16.6
1.25	6.2	6.6	7.0	7.4	7.8	8.3	8.8	9.3	9.8	10.3	10.9	11.5	12.0	12.6	13.2	13.8	14.4	15.0	15.6	16.2	16.8
1.5	6.7	7.0	7.4	7.8	8.2	8.6	9.1	9.6	10.1	10.6	11.2	11.7	12.3	12.9	13.4	14.0	14.6	15.2	15.8	16.4	17.0
1.75	7.1	7.5	7.8	8.2	8.6	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.6	13.1	13.7	14.3	14.8	15.4	16.0	16.6	17.2
2	7.6	7.9	8.3	8.6	9.0	9.5	9.9	10.4	10.8	11.3	11.8	12.4	12.9	13.4	14.0	14.6	15.1	15.7	16.3	16.8	17.4
2.25	8.1	8.4	8.7	9.1	9.5	9.9	10.3	10.8	11.2	11.7	12.2	12.7	13.3	13.8	14.3	14.9	15.4	16.0	16.6	17.1	17.7
2.5	8.6	8.9	9.2	9.6	10.0	10.4	10.8	11.2	11.7	12.1	12.6	13.1	13.6	14.2	14.7	15.2	15.8	16.3	16.9	17.5	18.0
2.75	9.1	9.4	9.7	10.1	10.4	10.8	11.2	11.7	12.1	12.6	13.1	13.5	14.0	14.6	15.1	15.6	16.2	16.7	17.3	17.8	18.4
3	9.6	9.9	10.2	10.6	10.9	11.3	11.7	12.1	12.6	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5	17.1	17.6	18.2	18.7
3.25	10.1	10.4	10.7	11.1	11.5	11.8	12.2	12.6	13.1	13.5	14.0	14.4	14.9	15.4	15.9	16.4	17.0	17.5	18.0	18.6	19.1
3.5	10.6	10.9	11.3	11.6	12.0	12.3	12.7	13.1	13.6	14.0	14.5	14.9	15.4	15.9	16.4	16.9	17.4	17.9	18.5	19.0	19.5
3.75	11.2	11.5	11.8	12.1	12.5	12.9	13.3	13.7	14.1	14.5	14.9	15.4	15.9	16.3	16.8	17.3	17.8	18.4	18.9	19.4	20.0
4	11.7	12.0	12.3	12.7	13.0	13.4	13.8	14.2	14.6	15.0	15.4	15.9	16.4	16.8	17.3	17.8	18.3	18.8	19.3	19.9	20.4
4.25	12.2	12.6	12.9	13.2	13.6	13.9	14.3	14.7	15.1	15.5	16.0	16.4	16.9	17.3	17.8	18.3	18.8	19.3	19.8	20.3	20.8
4.5	12.8	13.1	13.4	13.8	14.1	14.5	14.8	15.2	15.6	16.1	16.5	16.9	17.4	17.8	18.3	18.8	19.3	19.8	20.3	20.8	21.3
4.75	13.4	13.7	14.0	14.3	14.7	15.0	15.4	15.8	16.2	16.6	17.0	17.4	17.9	18.3	18.8	19.3	19.8	20.2	20.7	21.2	21.8
5	13.9	14.2	14.6	14.9	15.2	15.6	16.0	16.3	16.7	17.1	17.5	18.0	18.4	18.9	19.3	19.8	20.3	20.7	21.2	21.7	22.2

Table 7.15: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 2.5$ mm. Values of Σ_m and Σ_A are given in mm.

Σ_A \ Σ_m	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
0	5.2	5.8	6.3	6.9	7.5	8.1	8.7	9.3	9.9	10.5	11.1	11.7	12.3	12.9	13.5	14.1	14.7	15.4	16.0	16.6	17.2
0.25	5.5	5.9	6.4	7.0	7.5	8.1	8.7	9.3	9.9	10.5	11.1	11.7	12.3	12.9	13.5	14.2	14.8	15.4	16.0	16.6	17.2
0.5	5.9	6.3	6.7	7.2	7.7	8.3	8.8	9.4	10.0	10.6	11.2	11.8	12.4	13.0	13.6	14.2	14.8	15.4	16.0	16.7	17.3
0.75	6.3	6.6	7.0	7.5	8.0	8.5	9.0	9.6	10.2	10.7	11.3	11.9	12.5	13.1	13.7	14.3	14.9	15.5	16.1	16.7	17.3
1	6.6	7.0	7.4	7.8	8.3	8.8	9.3	9.8	10.4	10.9	11.5	12.1	12.7	13.2	13.8	14.4	15.0	15.6	16.2	16.8	17.4
1.25	7.1	7.4	7.8	8.2	8.6	9.1	9.6	10.1	10.6	11.2	11.7	12.3	12.8	13.4	14.0	14.6	15.2	15.8	16.4	17.0	17.6
1.5	7.5	7.8	8.2	8.6	9.0	9.5	9.9	10.4	10.9	11.4	12.0	12.5	13.1	13.6	14.2	14.8	15.4	16.0	16.5	17.1	17.7
1.75	7.9	8.3	8.6	9.0	9.4	9.8	10.3	10.8	11.3	11.8	12.3	12.8	13.4	13.9	14.5	15.0	15.6	16.2	16.8	17.3	17.9
2	8.4	8.7	9.0	9.4	9.8	10.2	10.7	11.1	11.6	12.1	12.6	13.1	13.7	14.2	14.7	15.3	15.9	16.4	17.0	17.6	18.2
2.25	8.8	9.2	9.5	9.9	10.3	10.7	11.1	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.1	15.6	16.2	16.7	17.3	17.9	18.4
2.5	9.3	9.6	10.0	10.3	10.7	11.1	11.5	12.0	12.4	12.9	13.4	13.9	14.4	14.9	15.4	15.9	16.5	17.0	17.6	18.2	18.7
2.75	9.8	10.1	10.4	10.8	11.2	11.6	12.0	12.4	12.8	13.3	13.8	14.3	14.7	15.3	15.8	16.3	16.8	17.4	17.9	18.5	19.0
3	10.3	10.6	10.9	11.3	11.6	12.0	12.4	12.9	13.3	13.7	14.2	14.7	15.2	15.7	16.2	16.7	17.2	17.7	18.3	18.8	19.4
3.25	10.8	11.1	11.4	11.8	12.1	12.5	12.9	13.3	13.7	14.2	14.6	15.1	15.6	16.1	16.6	17.1	17.6	18.1	18.7	19.2	19.7
3.5	11.3	11.6	11.9	12.3	12.6	13.0	13.4	13.8	14.2	14.7	15.1	15.6	16.0	16.5	17.0	17.5	18.0	18.5	19.1	19.6	20.1
3.75	11.8	12.1	12.4	12.8	13.1	13.5	13.9	14.3	14.7	15.1	15.6	16.0	16.5	17.0	17.4	17.9	18.4	19.0	19.5	20.0	20.5
4	12.3	12.6	13.0	13.3	13.7	14.0	14.4	14.8	15.2	15.6	16.1	16.5	17.0	17.4	17.9	18.4	18.9	19.4	19.9	20.4	20.9
4.25	12.9	13.2	13.5	13.8	14.2	14.5	14.9	15.3	15.7	16.1	16.5	17.0	17.4	17.9	18.4	18.9	19.3	19.8	20.3	20.9	21.4
4.5	13.4	13.7	14.0	14.4	14.7	15.1	15.4	15.8	16.2	16.6	17.1	17.5	17.9	18.4	18.8	19.3	19.8	20.3	20.8	21.3	21.8
4.75	13.9	14.2	14.6	14.9	15.2	15.6	16.0	16.3	16.7	17.1	17.6	18.0	18.4	18.9	19.3	19.8	20.3	20.8	21.3	21.8	22.3
5	14.5	14.8	15.1	15.4	15.8	16.1	16.5	16.9	17.3	17.7	18.1	18.5	18.9	19.4	19.8	20.3	20.8	21.3	21.7	22.2	22.7

Table 7.16: 3D margin calculated for $M_{m,i} = 0$ mm and $M_{A,i} = \pm 3.0$ mm. Values of Σ_m and Σ_A are given in mm.

Chapter 8

Summary

The presented work aimed to study an influence of patient setup errors on the radiotherapy process and distribution of received dose. Performed studies evaluated parametrization of population of patients' setup error, estimation of cumulative dose distribution, off-line verification protocols and CTV-PTV margins.

I started with analyzing two different clinical datasets of setup errors for prostate cancer treatments in Chapter 3. A group of 100 patients was treated in COI and their setup errors were defined on the basis of bony anatomy in 25 fractions treatment. The second group of patients was treated in Erasmus MC and consisted of 835 patients with 39 fractions treatment. The setup errors of these patients were acquired on the basis of gold fiducials implanted into the prostate. To the best of my knowledge this is the largest analyzed database for prostate cancer treatment. For both groups of patients the distribution of setup errors did not have a Gaussian distribution. That finding was in agreement with [51]. Evaluation was done to check whether this non-Gaussian distribution of setup errors may be attributed to inter-fraction time trends. As in radiotherapy there is always a finite number of fractions some trend can always be fitted to measured setup errors, even if there would be no trend in case of infinite fraction number. The test was provided to check probability that observed trends have a physiological and not statistical nature — see Section 3.6.

Many groups showed existence of inter-fraction time trends for different cancer sites [23, 29, 37, 67, 72, 83, 87, 93]. Therefore I decided to check ability of currently mostly applied conventional parametrization of population of patients' setup error to reproduce the original data (ground truth) and to reproduce behaviour of original data in case the offline verification protocol (NAL or eNAL) would be applied. The results showed that the conventional parametrization method overestimates performance of NAL off-line verification protocol for populations of patients experiencing inter-fraction time trends. The alternative trendline parametrization was also examined — see Chapter 4. It was shown that trendline parametrization can correctly describe population of patients experiencing time trends as well as their behaviour after application of verification protocols. Usage of a wrong parametrization for data evaluation may lead to overestimation of Σ_{res} and, in consequence, too small CTV-PTV margins. That is why it is of crucial importance to evaluate datasets in terms of inter-fraction setup errors existence and apply an appropriate parametrization. Currently investigated concepts of probability and margin-less planning are based on conventional parametrization [5, 27, 28]. Implementing a trendline parametrization for that purpose should be considered.

Currently used methods of CTV-PTV margin calculation [90] as well as methods used in probability and margin-less planning [5, 27, 28] are based on the cumulative dose distribution.

This cumulative dose distribution is estimated based on population parametrization. In Chapter 5 a novel method of estimating the cumulative patient dose distribution per patient was investigated. It has been shown that depending on RT technique, after first few fractions a patient specific estimation of cumulative dose distribution can be performed. That result can be used in several ways. First of all, it is quite easy to implement it to a clinical usage without the need of TPS change. Such implementation can improve methods of adaptive radiotherapy as a new plan can be generated on the basis of the estimated cumulative dose. Secondly, this approach can be also used in patient-specific probabilistic or margin-less planning.

In clinical practice it can be seen that for some patients undergoing radiotherapy an on-line verification is required even if it is not a part of procedure. Currently applied NAL protocol does not consider such patients. In NAL protocol all patients have exactly the same number of fractions with imaging. The quality of NAL performance depends on estimation of mean setup error done after first few fractions (usually 3 or 4). The new iNAL protocol was proposed in order to distinguish patients with potentially erroneous estimation of mean setup error (Chapter 6). In the iNAL protocol the number of fractions with imaging is individualized. Still, the average number of fractions with imaging can be as low as four. The proposed iNAL protocol was compared with NAL and iNAL protocols. The change in distribution of residual systematic errors showed that indeed problematic patients (large movers) can be distinguished. It was shown that with the same or even smaller margins, calculated on the basis of van Herk recipe, bigger group of large movers would have CTV within the margin.

Currently in clinical practice van Herk recipe is mostly used for CTV-PTV margin calculation. This recipe is based on the conventional parametrization of population of patients' setup error and therefore leads to too small margins for patients experiencing inter-fraction time trends. For big time trends even all the patients have at least one trendline error outside the margin. In chapter Chapter 7, the previously presented trendline parametrization was used to propose a new margin recipe. The proposed formalism of margin calculation guarantees that assumed number of patients outside the margin (usually 90%) is met. In case of no time trends the proposed recipe simplifies to van Herk's one. Three alternative methods of calculating margin with new recipe were presented: direct equation, Python code and look-up tables.

It can be concluded that aim of presented studies was reached. Analysis of large group of patients showed problems with currently applied methods. Solutions to overcome this problems were proposed and evaluated. Simple methods of radiotherapy individualization were also investigated. The proposed iNAL offline verification protocol can be applied without increase in workload. Presented method of patient-specific estimation of cumulative dose distribution does not require new TPS systems or algorithms.

Bibliography

[1] H Alasti, MP Petric, CN Catton, and PR Warde. Portal imaging for evaluation of daily on-line setup errors and off-line organ motion during conformal irradiation of carcinoma of the prostate. *International Journal of Radiation Oncology*Biology*Physics*, 49(3):869–884, 2001.

[2] F Alongi, A Fogliata, P Navarria, A Tozzi, P Mancosu, F Lobefalo, G Reggiori, A Clivio, L Cozzi, and M Scorsetti. Moderate Hypofraktionierung und simultan integrierter Boost mit volumetrisch modulierter Bogentherapie (RapidArc) bei Prostatakrebs. *Strahlentherapie und Onkologie*, 188(11):990–996, 2012.

[3] A Bel, M van Herk, H Bartelink, and JV Lebesque. A verification procedure to improve patient set-up accuracy using portal images. *Radiotherapy and Oncology*, 29(2):253–60, 1993.

[4] J Bijhold, JV Lebesque, AAM Hart, and RE Vijlbrief. Maximizing setup accuracy using portal images as applied to a conformal boost technique for prostatic cancer. *Radiotherapy and Oncology*, 24(4):261–271, 1992.

[5] R Bohoslavsky, MG Witte, TM Janssen, and M van Herk. Probabilistic objective functions for margin-less IMRT planning. *Physics in Medicine and Biology*, 58(11):3563–3580, 2013.

[6] T Bortfeld, S Jiang, and E Rietzel. Effects of motion on the total dose distribution. *Seminars in Radiation Oncology*, 14(1):41–51, 2004.

[7] T Bortfeld, M van Herk, and SB Jiang. When should systematic patient positioning errors in radiotherapy be corrected? *Physics in Medicine and Biology*, 47(28):N297–302, 2002.

[8] T Bortfeld and S Webb. Single-Arc IMRT? *Physics in Medicine and Biology*, 54(1):N9–N20, 2008.

[9] NG Burnet, SJ Thomas, KE Burton, and SJ Jefferies. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging*, 4(2):153–161, 2004.

[10] Z Chen, Z Yang, J Wang, and W Hu. Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy. *Radiation Oncology*, 11(103), 2016.

[11] T Craig, J Battista, and J van Dyk. Limitations of a convolution method for modeling geometric uncertainties in radiation therapy. I. the effect of shift invariance. *Medical Physics*, 30(8):2001–2011, 2003.

[12] T Craig, J Battista, and J van Dyk. Limitations of a convolution method for modeling geometric uncertainties in radiation therapy. II. the effect of a finite number of fractions. *Medical Physics*, 30(8):2012–2020, 2003.

[13] LA Dawson, K Mah, E Franssen, and G Morton. Target position variability throughout prostate radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 42(5):1155–1161, 1998.

[14] HCJ de Boer and BJM Heijmen. A protocol for the reduction of systematic patient setup errors with minimal portal imaging workload. *International Journal of Radiation Oncology*Biology*Physics*, 50(5):1350–1365, 2001.

[15] HCJ de Boer and BJM Heijmen. A new approach to off-line setup corrections: Combining safety with minimum workload. *Medical Physics*, 29(9):1998–2012, 2002.

[16] HCJ de Boer and BJM Heijmen. eNAL: An Extension of the NAL Setup Correction Protocol for Effective Use of Weekly Follow-up Measurements. *International Journal of Radiation Oncology*Biology*Physics*, 67(5):1586–1595, 2007.

[17] HCJ de Boer, MJ van Os, PP Jansen, and BJM Heijmen. Application of the No Action Level (NAL) protocol to correct for prostate motion based on electronic portal imaging of implanted markers. *International Journal of Radiation Oncology*Biology*Physics*, 61(4):969–983, 2005.

[18] P DeLuca, D Jones, R Gahbauer, G Whitmore, and A ÌĄWambersie. *ICRU report 83: Prescribing, Recording, and Reporting Photon-BeamIntensity-Modulated Radiation Therapy (IMRT)*. International Commission on Radiation Units and Measurements, 2010.

[19] T Depuydt, A van Esch, and DP Huyskens. A quantitative evaluation of IMRT dose distributions: refinement and clinical assessment of the gamma evaluation. *Radiotherapy and Oncology*, 62(3):309–319, 2002.

[20] J Didkowska and U Wojciechowska. Nowotwory w Polsce w 2012 roku. *Nowotwory. Journal of Oncology*, 63(3):197–216, 2013.

[21] M Dolezel, K Odrazka, M Vaculikova, J Vanasek, J Sefrova, P Paluska, M Zouhar, J Jansa, Z Macingova, L Jarosova, M Brodak, P Moravek, and I Hartmann. Dosissteigerung bei Prostatabestrahlung durch einen simultanen integrierten Boost bis 82 Gy. Vergleich der Akut und Spättoxizität nach 3D-CRT bis 74 Gy und IMRT bis 78 Gy. *Strahlentherapie und Onkologie*, 186(4):197–202, 2010.

[22] Y Donner, K Fortney, SRG Calimport, K Pfleger, M Shah, and J Betts-LaCroix. Great Desire for Extended Life and Health amongst the American Public. *Frontiers in Genetics*, 6, 2016.

[23] Ali A H El-Gayed, Arjan Bel, Ron Vijlbrief, Harry Bartelink, and Joos V Lebesque. Time trend of patient setup deviations during pelvic irradiation using electronic portal imaging. *Radiotherapy and Oncology*, 26(2):162–171, February 1993.

[24] C Fiorino, C Cozzarini, and P Passoni. The promise of adaptive radiotherapy for pelvic tumors: "too high cost for too little result" or "a low cost for a significant result"? *Acta Oncologica*, 55(8):939–942, 2016.

[25] C Fiorino, M Reni, A Bolognesi, GM Cattaneo, and R Calandrino. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and Oncology*, 47(3):285–292, 1998.

[26] European Organisation for Research and Treatment of Cancer. EORTC Guidelines. `https://www.eortc.org/guidelines/`. Accessed: 2019-03-15.

[27] A Fredriksson and R Bokrantz. The scenario-based generalization of radiation therapy margins. *Physics in Medicine and Biology*, 61(5):2067–2082, 2016.

[28] A Fredriksson, A Forsgren, and B Hårdemark. Maximizing the probability of satisfying the clinical goals in radiation therapy treatment planning under setup uncertainty. *Medical Physics*, 42(7):3992–3999, 2015.

[29] A Gangsaas, E Astreinidou, S Quint, PC Levendag, and B Heijmen. Cone-Beam Computed Tomography - guided positioning of laryngeal cancer patients with large interfraction time trends in setup and nonrigid anatomy variations. *International Journal of Radiation Oncology*Biology*Physics*, 87(2):401–406, 2013.

[30] A Ghasemi and S Zahediasl. Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology & Metabolism*, 10(2):486–489, 2012.

[31] S Gill, D Pham, K Dang, M Bressel, T Kron, S Siva, PK Tran, KH Tai, and F Foroudi. Plan of the day selection for online image-guided adaptive post-prostatectomy radiotherapy. *Radiotherapy and Oncology*, 107(2):165–170, 2013.

[32] M Giżyńska. Analiza rozkładów dawek w radioterapii z zastosowaniem modulacji intensywności dawki w porównaniu z radioterapią konformalną. Master's thesis, University of Warsaw, 2006.

[33] M Giżyńska, D Blatkiewicz, B Czyżew, M Gałecki, M Gil-Ulkowska, and P Kukołowicz. Prediction of the cumulated dose for external beam irradiation of prostate cancer patients with 3D-CRT technique. *Nukleonika*, 61(1):15–18, 2016.

[34] M Giżyńska, P Kukołowicz, and BJ Heijmen. Coping with inter-fraction time trends in tumor set-up. *Medical Physics*, submitted.

[35] EJ Hall and CS Wuu. Radiation-induced second cancers: the impact of 3D-CRT and IMRT. *International Journal of Radiation Oncology*Biology*Physics*, 56(1):83–88, 2003.

[36] JA Hanley, JA Hanley, and BJ McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.

[37] Joseph Hanley, Moira A Lumley, Gig S Mageras, Jerry Sun, Michael J Zelefsky, Steven A Leibel, Zvi Fuks, and Gerald J Kutcher. Measurement of patient positioning errors in three-dimensional conformal radiotherapy of the prostate. *International Journal of Radiation Oncology*Biology*Physics*, 37(2):435–444, January 1997.

[38] Z Hanusz, NM Razali, J Tarasińska, and YB Wah. Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality. *Biometrical Letters*, 52(2):85–93, 2011.

[39] A Heidenreich, P J Bastian, J Bellmunt, and M Bolla. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent—update 2013. *European Urology*, 65(1):124–137, 2014.

[40] A Heidenreich, J Bellmunt, M Bolla, S Joniau, and M Mason. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. *European Urology*, 59(1):61–71, 2011.

[41] ST Heijkoop, TR Langerak, S Quint, L Bondar, JW Mens, and MS Heijmen, BJ adn Hoogeman. Clinical implementation of an online adaptive plan-of-the-day protocol for nonrigid motion management in locally advanced cervical cancer IMRT. *International Journal of Radiation Oncology*Biology*Physics*, 90(3):673–679, 2014.

[42] CW Hurkmans, P Remeijer, JV Lebesque, and BJ Mijnheer. Set-up verification using portal imaging; review of current clinical practice. *Radiotherapy and Oncology*, 58(2):105–120, 2001.

[43] ML Johnston, P Vial, KL Wiltshire, LJ Bell, S Blome, Z Kerestes, GW Morgan, D O'Driscoll, TP Shakespeare, and TN Eade. Daily Online Bony Correction is Required for Prostate Patients Without Fiducial Markers or Soft-tissue Imaging. *Clinical Oncology*, 23(7):454–459, 2011.

[44] Polskie Towarzystwo Onkologii Klinicznej. Zalecenia postępowania diagnostyczno-terapeutycznego w nowotworach złośliwych 2019 rok. `http://onkologia.zalecenia.med.pl/`. Accessed: 2019-03-15.

[45] KD Kochanek, SL Murphy, and J Xu. Deaths: final data for 2014. *National Vital Statistics Reports*, 65(4):1–122, 2016.

[46] P Kukołowicz. The individualization in radiotherapy: Where are we now, and where are we going? *Physica Medica*, 32(S3):169, 2016.

[47] T Landberg, J Chavaudra, J Dobbs, G Hanks, K-A Johansson, T MÃűller, and J Purdy. *ICRU report 50: Prescribing, recording and reporting photon beam therapy*. International Commission on Radiation Units and Measurements, 1993.

[48] JV Lebesque and RB Keus. The simultaneous boost technique: the concept of relative normalized total dose. *Radiotherapy and Oncology*, 22(1):45–55, 1991.

[49] XA Li, JZ Wang, PA Jursinic, CA Lawton, and D Wang. Dosimetric advantages of IMRT simultaneous integrated boost for high-risk prostate cancer. *International Journal of Radiation Oncology*Biology*Physics*, 61(4):1251–1257, 2005.

[50] C Lin, S Xu, W Yao, Y Wu, J Fang, and VWC Wu. Comparison of set up accuracy among three common immobilization systems for intensity modulated radiotherapy of nasopharyngeal carcinoma patients. *Journal of Medical Radiation Sciences*, 64(2):106–113, 2017.

[51] Y Lin, T Liu, W Yang, X Yang, and MK Khan. The non-Gaussian nature of prostate motion based on real-time intrafraction tracking. *International Journal of Radiation Oncology*Biology*Physics*, 87(2):363–369, 2013.

[52] JM Links, LS 2nd Beach, B Subramaniam, MA Rubin, JG Hennessey, and AL Reiss. Edge complexity and partial volume effects. *Journal of computer assisted tomography*, 22(3):450–458, 1998.

[53] DA Low, WB Harms, S Mutic, and JA Purdy. A technique for the quantitative evaluation of dose distributions. *Medical Physics*, 25(5):656–661, 1998.

[54] EM Lozano, LA Pérez, J Torres, C Carrascosa, M Sanz, F Mendicote, and A Gil. Correction of systematic set-up error in breast and head and neck irradiation through a no-action level (NAL) protocol. *Clinical and Translational Oncology*, 13(1):34–42, 2011.

[55] D Martens, M Luesink, H Huizenga, and KL Pasma. eNAL++: a new and effective offline correction protocol for rotational setup errors when using a robotic couch. *Journal of Applied Clinical Medical Physics*, 16(6):177–185, 2015.

[56] L Martins, JG Couto, and B Barbosa. Use of planar kv vs. cbct in evaluation of setup errors in oesophagus carcinoma radiotherapy. *Reports of Practical Oncology & Radiotherapy*, 21(1):57–62, 2016.

[57] HA McNair, VN Hansen, CC Parker, Phil M Evans, A Norman, E Miles, EJ Harris, L DeLacroix, E Smith, R Keane, VS Khoo, AC Thompson, and DP Dearnaley. A Comparison of the Use of Bony Anatomy and Internal Markers for Offline Verification and an Evaluation of the Potential Benefit of Online and Offline Verification Protocols for Prostate Radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 71(1):41–50, 2007.

[58] GJ. Meijer, C Rasch, P Remeijer, and JV Lebesque. Three-dimensional analysis of delineation errors, setup errors, and organ motion during radiotherapy of bladder cancer. *International Journal of Radiation Oncology*Biology*Physics*, 55(5):1277–1287, 2003.

[59] M Miften, A Olch, D Mihailidis, and J Moran. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Medical Physics*, 5(4):e53–e83, 2018.

[60] B Mijnheer, A Olszewska, C Fiorino, and G Hartmann. *Quality assurance of treatment planning systems: practical examples for non-IMRT photon beams*. European Society for Radiotheraphy and Oncology, 2004.

[61] MR Moman, UA van der Heide, ANTJ Kotte, RJA van Moorselaar, GH Bol, SPG Franken, and M van Vulpen. Long-term experience with transrectal and transperineal implantations of fiducial gold markers in the prostate for position verification in external beam radiotherapy; feasibility, toxicity and quality of life. *Radiotherapy and Oncology*, 96(1):38–42, 2010.

[62] JA Moore, JJ Gordon, M Anscher, J Silva, and JV Siebers. Comparisons of treatment optimization directly incorporating systematic patient setup uncertainty with a margin-based approach. *Medical Physics*, 39(2):1102–1111, 2012.

[63] JA Moore, JJ Gordon, MS Anscher, and Jeffrey V Siebers. Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach. *Medical Physics*, 36(9):3880–3890, 2009.

[64] K Moore, C Paterson, J Hicks, S Harrow, and M McJury. Stereotactic ablative body radiotherapy for non-small-cell lung cancer: setup reproducibility with novel arms-down immobilization. *The British Journal of Radiology*, 89(1068), 2016.

[65] AJ Mundt and JC Roeske. *Intensity Modulated Radiation Therapy*. BC Decker Inc, 2005.

[66] CA Nalder, AM Bidmead, CD Mubata, D Tait, and C Beardmore. Influence of a vacfix immobilization device on the accuracy of patient positioning during routine breast radiotherapy. *The British Journal of Radiology*, 74(879):249–254, 2001.

[67] A Namysł-Kaletka, J Wydmanski, A Tukiendorf, D Bodusz, W Leszczynski, R Kawczynski, K Grabinska, and P Polanowski. Influence of interfraction motion on margins for radiotherapy of gastric cancer. *The British Journal of Radiology*, 88(1048), 2015.

[68] H Nightingale, R Conroy, T Elliott, C Coyle, JP Wylie, and A Choudhury. A national survey of current practices of preparation and management of radical prostate radiotherapy patients during treatment. *Radiography*, 23(2):87–93, 2017.

[69] American Society of Clinical Oncology. ASCO Guidelines by Clinical Area. `https://www.asco.org/practice-guidelines/quality-guidelines/guidelines`. Accessed: 2019-03-15.

[70] K Otto. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Medical Physics*, 35(1):310–317, 2008.

[71] D Palma, E Vollans, K James, S Nakano, V Moiseenko, R Shaffer, M McKenzie, J Morris, and K Otto. Volumetric modulated arc therapy for delivery of prostate radiotherapy: comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 72(4):996–1001, 2008.

[72] J Penninkhof, S Quint, M Baaijens, B Heijmen, and M Dirkx. Practical use of the extended no action level (enal) correction protocol for breast cancer patients with implanted surgical clips. *International Journal of Radiation Oncology*Biology*Physics*, 82(2):1031–1037, 2012.

[73] M Pinkawa, B Asadpour, B Gagel, and MD Piroth. Prostate position variability and dose–volume histograms in radiotherapy for prostate cancer with full and empty bladder. *International Journal of Radiation Oncology*Biology*Physics*, 64(3):856–861, 2006.

[74] M Piziorska, P Kukołowicz, A Zawadzka, M Pilichowska, and P Pęczkowski. Adaptives Offline-Protokoll mit digitaler Volumentomographie für die perkutane Strahlentherapie des Prostatakarzinoms. *Strahlentherapie und Onkologie*, 188(11):1003–1009, 2012.

[75] MMG Rossi, HMU Peulen, JSA Belderbos, and J-J Sonke. Intrafraction motion in stereotactic body radiation therapy for non-small cell lung cancer: Intensity modulated radiation

therapy versus volumetric modulated arc therapy. *International Journal of Radiation Oncology*Biology*Physics*, 95(2):835–843, 2016.

[76] M Rybovic, RB Banati, and J Cox. Radiation therapy treatment verification imaging in Australia and New Zealand. *Journal of Medical Imaging and Radiation Oncology*, 52(2):183–190, 2008.

[77] S Sawayanagi, H Yamashita, M Ogita, T Kiritoshi, T Nakamoto, O Abe, and K Nakagawa. Volumetric and dosimetric comparison of organs at risk between the prone and supine positions in postoperative radiotherapy for prostate cancer. *Radiation Oncology*, 13(70), 2018.

[78] W Schlegel and A Mahr. *3D conformal radiation therapy: Multimedia introduction to methods and techniques.* Springer Publishing Company, Incorporated, 2007.

[79] SS Shapiro and MB Wilk. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4):591–611, 1965.

[80] R Siegel and D Naishadham. Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, 63(1):11–30, 2013.

[81] RJHM. Steenbakkers, JC Duppen, A Betgen, HT Lotz, P Remeijer, I Fitton, PJCM Nowak, M van Herk, and CRN Rasch. Impact of knee support and shape of tabletop on rectum and prostate position. *International Journal of Radiation Oncology*Biology*Physics*, 60(5):1364–1372, 2004.

[82] JC Stroom, HCJ de Boer, and H Huizenga. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *International Journal of Radiation Oncology*Biology*Physics*, 43(4):905–919, 1999.

[83] JC Stroom, PCM Koper, GA Korevaaar, M van Os, M Janssen, HCJ de Boer, PC Levendag, and BJM Heijmen. Internal organ motion in prostate cancer patients treated in prone and supine treatment position. *Radiotherapy and Oncology*, 51(3):237–248, 1999.

[84] LA Torre, F Bray, RL Siegel, and J Ferlay. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65(2):87–108, 2015.

[85] HS Tsang, CP Kamerling, P Ziegenhein, S Nill, and U Oelfke. A novel probabilistic approach to generating PTV with partial voxel contributions. *Physics in Medicine and Biology*, 62(12):4917–4928, 2017.

[86] AH Udayashankar, S Noorjahan, N Srikantia, KR Babu, and S Muzumder. Immobilization versus no immobilization for pelvic external beam radiotherapy. *Reports of Practical Oncology & Radiotherapy*, 23(4):233–241, 2018.

[87] UA van der Heide, ANTJ Kotte, H Dehnad, P Hofman, JJW Lagenijk, and M van Vulpen. Analysis of fiducial marker-based position verification in the external beam radiotherapy of patients with prostate cancer. *Radiotherapy and Oncology*, 82(1):38–45, 2007.

[88] AR van Erkel and PM Pattynama. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *European Journal of Radiology*, 27(2):88–94, 1998.

[89] M van Herk. Errors and margins in radiotherapy. *Seminars in radiation oncology*, 14(1):52–64, 2004.

[90] M van Herk, P Remeijer, C Rasch, and JV Lebesque. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 47(4):1121–1135, 2000.

[91] M van Herk, M Witte, and P Remeijer. Performance of patient specific margins derived using a bayesian statistical method. In *IFMBE Proceedings*, pages 769–771, 2009.

[92] A Wambersie and T Landgerg. *ICRU report 62: Prescribing, recording and reporting photon beam therapy*. International Commission on Radiation Units and Measurements, 1999.

[93] E Weiss, SP Robertson, N Mukhopadhyay, and GD Hugo. Tumor, Lymph node and Lymph Node-to-Tumor Displacements over a Radiotherapy Series: Analysis of Inter- and Intrafraction Variations using Active Breathing Control (ABC) in Lung Cancer. *International Journal of Radiation Oncology*Biology*Physics*, 82(4):e639–e645, 2002.

[94] J Wu, T Haycocks, H Alasti, G Ottewell, N Middlemiss, M Abdolell, P Warde, A Toi, and C Catton. Positioning errors and prostate motion during conformal prostate radiotherapy using on-line isocentre set-up verification and implanted prostate markers. *Radiotherapy and Oncology*, 61(2):127–133, 2001.

[95] D Yan, D Lockman, D Brabbins, L Tyburski, and A Martinez. An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *International Journal of Radiation Oncology*Biology*Physics*, 48(1):289–302, 2000.

[96] D Yan, E Ziaja, D Jaffray, J Wong, and D Brabbins. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study. *International Journal of Radiation Oncology*Biology*Physics*, 41(3):715–720, 1998.

[97] C X Yu. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Physics in Medicine and Biology*, 40(9):1435–1449, 1999.

[98] MJ Zelefsky, Z Fuks, L Happersett, HJ Lee, CC Ling, CM Burman, M Hunt, T Wolfe, ES Venkatraman, A Jackson, M Skwarchuk, and SA Leibel. Clinical experience with intensity modulated radiation therapy (IMRT) in prostate cancer. *Radiotherapy and Oncology*, 55(3):241–249, 2000.